

UNIVERSITÉ DE MONTRÉAL

**MODÈLE PROBABILISTE DE COURBES DÉFORMABLES
POUR APPARIEMENT AVEC IMAGES DE SCÈNE SEMITRANSARENTE**

**HAIL MALLOUCHE
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL**

**THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIAE DOCTOR (Ph.D.)
(GÉNIE BIOMÉDICAL)
JUN 1997**



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-33012-5

UNIVERSITÉ DE MONTRÉAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

MODÈLE PROBABILISTE DE COURBES DÉFORMABLES
POUR APPARIEMENT AVEC IMAGES DE SCÈNE SEMITRANSARENTE

présentée par: MALLOUCHE Hail

en vue de l'obtention du diplôme de: Philosophiae Doctor (Ph.D.)

a été dûment acceptée par le jury d'examen constitué de:

M. DANSEREAU Jean, Ph.D., président

M. GOUSSARD Yves, Ph.D., membre et directeur de recherche

M. DE GUISE Jacques A., Ph.D., membre et co-directeur de recherche

M. MEUNIER Jean, Ph.D., membre

M. MITICHE Amar, Ph.D., membre externe

*« En vérité,
la distance la plus longue est celle qui s'étale entre
la vision de votre sommeil et celle de votre réveil,
entre ce qui n'est qu'action et ce qui est désir. »*

Gibran Khalil GIBRAN

*... une vision
d'espoir et de liberté
pour ceux qui en manquent ...*

Remerciements

Je tiens d'abord à remercier les professeurs Yves Goussard et Jacques A. de Guise, mes deux directeurs, pour leur soutien, leur patience et leur disponibilité tout au long de ce travail.

J'aimerais également exprimer mes remerciements à l'Institut de génie biomédical et à son directeur le professeur Robert A. Leblanc pour m'avoir si bien accueilli dans un environnement humain, professionnel et compétent pendant mes années aux études supérieures. Je mentionne aussi le professeur Pierre Mathieu pour l'attention particulière qu'il a prêtée aux dossiers académiques. Je tiens, de plus, à saluer Mesdames Diane Giroux et Louise Clément pour leurs compréhension, gentillesse et excellent travail.

Mes remerciements s'adressent également au Laboratoire d'imagerie, de vision, et d'intelligence artificielle de l'École de technologie supérieure (LIVIA) de l'Université du Québec, qui m'a offert le support nécessaire au développement de cette recherche. Je remercie aussi Mme Danielle Boutlot et M. Yves Nadeau du personnel technique du laboratoire pour leur excellent support.

J'aimerais aussi souligner la collaboration du centre de recherche sur l'étude de la scoliose de l'Hôpital Sainte-Justine de Montréal. En particulier, je remercie Dr. Hubert Labelle pour le support clinique et M. André Beauchamp pour le support technique. De plus, j'aimerais souligner les importantes réunions du groupe de recherche sous la direction du professeur Jean Dansereau du département de génie mécanique de l'École Polytechnique.

J'adresse aussi mes remerciements les plus sincères à toute l'École Polytechnique de Montréal pour son excellent milieu académique et de recherche. Je n'oublierai pas l'aimable Mme Louise Grenon du service des études supérieures pour sa gentillesse et générosité.

Je ne voudrais pas oublier de remercier les différents organismes et personnes pour leurs bourses et subventions sans lesquelles cette recherche n'aurait pas été possible: (i) Les professeurs Jacques A. de Guise et Yves Goussard pour leur soutien financier; (ii) Le conseil de recherche en sciences naturelles et en génie (CRSNG) du Canada pour une bourse doctorale; (iii) Les Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR) du Québec pour une bourse doctorale; (iv) Le conseil canadien des ingénieurs représenté par M. Daniel Verreault (président), et la compagnie Nord-Américaine d'assurance-vie représentée par Mme Grace Cohen (vice-présidente), pour une bourse d'excellence aux études supérieures; (v) Le centre international des étudiants et stagiaires de France pour une bourse de stage doctoral au laboratoire des signaux et systèmes (LSS) du centre national de la recherche scientifique à l'École supérieure d'électricité (Gif-Sur-Yvette, Paris). J'aimerais remercier le Dr. Jérôme Idier du LSS pour des échanges fort intéressants lors de mon passage au LSS.

Je remercie finalement tous les membres du jury dont les commentaires et les corrections ont aidé à rehausser la qualité de la thèse.

Résumé

Un système de vision artificielle peut avoir différents objectifs comme, par exemple, la segmentation et l'étiquetage des régions, la reconnaissance et la poursuite des formes en mouvement et la compréhension et la reconstruction 3-D des scènes à partir d'un nombre limité d'images 2-D. En imagerie médicale, les applications de la vision artificielle sont nombreuses, comme par exemple la délimitation des cavités cardiaques, l'estimation de la contraction du myocarde et la reconstruction 3-D du réseau coronarien et de la colonne vertébrale.

L'objectif de cette thèse est de développer un système de vision pour la détection et l'étiquetage de structures osseuses dans des images radiographiques. Ces images sont généralement bruitées et complexes à cause respectivement des processus de diffusion photonique et de la superposition semitransparente des structures. De plus, la structure des organes humains est complexe, non polyédrique et non rigide. Tous ces éléments rendent ce problème inverse mal-posé et par conséquent difficile à résoudre.

Jusqu'à maintenant, les principaux travaux de détection dans les images médicales se sont contentés de proposer des modèles libres pour analyser des formes simples, ou des modèles élastiques pour analyser des formes plus complexes. Cependant, la performance des modèles élastiques est limitée par le niveau du bruit et par l'amplitude de la disparité entre le modèle et l'objet réel dans l'image.

Nous proposons, dans cette thèse, une approche originale basée sur l'appariement d'un modèle déformable avec l'image observée. On formule le problème dans un cadre stochastique permettant de trouver une solution globale en utilisant un seul critère de recherche. En effet, on propose le critère de la maximisation de la *pseudo-probabilité conjointe* (MPPC) car il intègre des propriétés intéressantes: sa forme distributive et la dépendance locale de la solution par rapport à l'image observée et au modèle connu *a priori*. Pour mieux représenter la semitransparence des projections radiographiques et la géométrie non rigide des organes, on propose un modèle libre et déformable constitué de la superposition d'un ensemble de simples courbes fermées (SCFs). La continuité et la fermeture des SCFs aideront à réduire la dimension de l'espace de recherche. L'approche est développée d'abord avec un modèle *simple* formé uniquement de courbes extrémales et ensuite, avec un modèle plus *réaliste* formé de courbes extrémales et de courbes internes. Ce dernier modèle permet de représenter les projections semitransparentes d'objets volumétriques. La forme finale du critère de recherche exploite la dualité ligne-région des SCFs afin d'assurer la robustesse, la précision et la cohérence de détection tout en permettant d'introduire une interaction locale sans biais de niveau supérieur entre la solution et les informations topologiques du modèle. Cette interaction stabilise le processus de recherche de solution et aide à éviter plusieurs optima locaux.

Nous utilisons un processus de recherche de solution qui est déterministe et constitué de deux étapes: l'alignement du modèle à l'image par transformations globales; et l'appariement du modèle avec l'image par déformation aléatoire. Notre nouvelle technique d'alignement global est basée sur l'intensité de l'image et sur un critère de maximisation de *vraisemblance*.

La technique est itérative et utilise un modèle dynamique de transformations linéaires à action hiérarchique. Malgré le fait que les méthodes d'alignement basées sur l'intensité sont généralement sensibles au bruit, notre technique est stable et robuste au bruit et au processus de formation d'image. Elle ne dépend pas d'une étape préalable d'extraction de primitives et converge même dans des conditions de grande disparité entre le modèle et la solution.

La technique d'appariement est itérative et déforme progressivement et de façon aléatoire le modèle jusqu'à atteindre une solution optimale. Chaque déformation acceptée doit optimiser la fonction d'objectif développée à partir du critère de la MPPC. Cette fonction est donnée par la somme de trois termes: deux termes qui mesurent respectivement l'adéquation entre la solution et l'image observée et l'adéquation entre la solution et les connaissances topologiques du modèle; et un terme de lissage. De plus, la déformation acceptée doit vérifier des contraintes algorithmiques assurant la préservation des propriétés de simple courbe fermée (CSF) ainsi que l'aspect topologique de la solution.

Afin d'accélérer la convergence et d'améliorer la possibilité de surmonter les optima locaux, la technique d'appariement est intégrée dans une structure pyramidale de multirésolution qui augmente la distance d'interaction entre le modèle et l'image.

L'approche est implantée et expérimentée à l'aide d'images synthétiques et d'images réelles d'objets synthétiques. Les résultats d'appariement sont concluants. La détection des discontinuités dans les images et l'étiquetage des régions sont précis, robustes, stables et cohérents, même avec des images fortement bruitées à contraste faible et variable, ou présentant une grande disparité par rapport au modèle.

Les contributions majeures de cette thèse se résument par: une nouvelle technique

d'alignement global basée sur l'intensité et sur la maximisation de la distribution de *vraisemblance*; un cadre général d'analyse de scènes semitransparentes utilisant un nouveau modèle de frontières formé de la superposition de multiples CSFs; une nouvelle approche de détection et d'étiquetage par appariement du modèle libre maximisant une distribution de *pseudo-probabilité conjointe*; l'intégration, dans le critère de MPPC, d'une fonction d'interaction entre la solution et les connaissances topologiques du modèle sans affecter la précision de la détection; et une technique d'appariement par déformation aléatoire incorporant des contraintes algorithmiques pour préserver l'aspect topologique de la solution.

Abstract

Computer vision research aims at detection, labeling, object recognition, 3-D scene understanding and reconstruction from a small number of images. Computer vision research has many applications in medical imaging such as, e.g., delineation of endocardiac cavities, myocardial motion estimation and 3-D reconstruction of the coronary tree, of the spine and of the rib cage.

The goal of this thesis is to develop a machine vision system for delineating and labeling objects in images of semitransparent scenes. The specific application of the system is the detection of bone structures in X-ray medical images. These images are noisy and complex mainly because of the photon image formation process and the semitransparent superposition of the structures. Moreover, the shape of human organs is complex, non polyhedral and non rigid. Hence, this visual inverse problem is heavily ill-posed and difficult to solve. Heretofore, proposals for shape detection in medical images were centered on simple models for images of non complex objects and on elastic models for more complex structures. However, the precision of elastic methods can be heavily compromised when applied to noisy images presenting a high disparity with respect to the model.

In this thesis, we propose a new approach to object detection and labeling in X-ray images. The method is based on matching a deformable model with the observed image. The matching problem is formulated in a stochastic framework. This allows us to explore the

search space using a single global criterion. Thus, we propose the maximum *pseudo-joint* criterion (MPJ) because it integrates interesting properties: the distributive form and the local dependency of the solution with respect to the observed image and to the *a priori* known model. Deformable boundary models are convenient for matching of non rigid objects with images from semitransparent scenes. Our model is formed of the superposition of a set of free form deformable simple closed curves (SCCs). Continuity and closure properties of SCCs implicitly help reducing the search space dimension. The approach is first developed for a *simple* model using extremal curves only, and then for a *complete* model that incorporates additional inner curves. The *complete* model is more realistic and can represent the semitransparent projection of volumetric objects. Moreover, the natural region-line duality of SCCs is used to ensure high robustness, precision and coherence to the final objective function and to introduce a local unbiased high-level interaction between the solution and the topological *a priori* knowledge of the object.

We develop a two-step descent search process: model alignment by global transformations and model-image matching by progressive deformation. The new alignment technique is intensity-based and seeks to maximize a *likelihood* estimator. The technique is iterative and uses a set of dynamic hierarchical linear transformations. Unlike other intensity-based methods, our technique is stable and robust with respect to the noise and the image formation process. As opposed to feature-based methods, our technique does not require a feature extraction operation and performs well even under large disparity conditions.

The progressive matching technique randomly deforms the model in order to reach an optimal solution. Each retained deformation fulfills functional and algorithmic constraints. The

functional constraint translates the MPJ criterion whose objective function is given by the sum of two solution consistency terms with the observed image and the topological knowledge of the model, respectively, and a smoothing term. The algorithmic constraints ensure that deformations preserve the SCC properties and the topological aspect of the solution.

In order to speed up the convergence toward the solution and help overcome many local optima, the deformation technique is integrated in a pyramidal multiresolution structure that increases the range of interaction between the model and the image.

The approach is implemented and successfully tested with synthetic images and with X-ray images of a synthetic object. The system is capable of precise and robust detection, and of consistent and stable labeling of extremely noisy images with weak and variable contrast under large disparity conditions.

The major contributions of the thesis are the following: a new intensity-based technique for global alignment by maximization of a *likelihood* distribution; a new framework for semitransparent scenes analysis with a new boundary model of multiple superposed SCCs; a new approach for detection and labeling based on matching free form models by maximizing the *pseudo-joint* distribution; the integration in the MPJ criterion of an interaction function between the solution and the topologic knowledge of the model, without compromising the precision of detection; and a new matching technique by progressive and random deformation with algorithmic constraints for preserving the topological aspect of the solution.

Table de matières

Dédicace	iv
Remerciements	v
Résumé	vii
Abstract	xi
Table de matières	xiv
Liste des tableaux	xviii
Liste des figures	xix
Liste des symboles	xxiii
Liste des annexes	xxviii
Préface	xxix
Chapitre 1	
Introduction	1
1.1 Problématique générale	3
1.1.1 Vision artificielle	4
1.1.2 Le problème inverse	6
1.1.3 L'espace de recherche	9
1.1.4 Organisation perceptuelle	12
1.2 Problématique particulière à l'analyse des scènes semitransparentes	14
1.2.1 Difficultés reliées au processus de formation d'image	15
1.2.2 Difficultés reliées à la nature des scènes	17
1.3 Objectifs et hypothèses	19
1.4 La thèse	22
Chapitre 2	
Présentation générale de l'approche	25

2.1	Introduction	25
2.2	Modèle géométrique	28
2.2.1	Revue des modèles géométriques	28
2.2.2	Modèle de multiples simples courbes fermées	31
2.3	Alignement global du modèle	32
2.3.1	Revue des méthodes d'alignement global	32
2.3.2	Modèle dynamique pour alignement global	33
2.4	Appariement du modèle par déformation	34
2.4.1	Revue des méthodes de recherche	35
2.4.2	Appariement par déformation aléatoire	35
2.4.3	Critère probabiliste	36
2.4.4	Contraintes algorithmiques	37
2.5	Conclusion	38

Chapitre 3

Alignement global du modèle

Article: *Dynamic model for global alignment and scaling* 40

3.0	Résumé	40
3.1	Introduction	45
3.2	Global alignment methods	47
3.3	Method	51
3.3.1	The stochastic model	53
3.3.2	The dynamic model	56
3.3.3	The resultant forces	61
3.3.4	Iterative time integration	62
3.3.5	Hierarchical transformations	64
3.3.6	Extensions	66
3.4	Implementation and experimental results	67
3.4.1	Implementation	68
3.4.2	Results	70
3.5	Conclusion	72

Chapitre 4

Modèle probabiliste d'appariement

Article: *Probabilistic model of multiple dynamic curve matching for semitransparent scene of nonrigid objects* 82

4.0	Résumé	82
4.1	Introduction	87
4.2	Problem Statement	89
4.3	Related Work	90
4.4	Matching Multiple Dynamic Curves	93
4.4.1	Maximum Pseudo-Joint Criterion	95
4.4.2	Model for Multiple Dynamic Curves	97

4.5	Data Driven Term	100
4.5.1	General Form	101
4.5.2	Local Data Driven Term - Discussion	105
4.6	Model Driven Term	107
4.7	Objective Function	110
4.8	Implementation and Optimization Method	111
4.8.1	Object Model	111
4.8.2	Explicit Forms of f_k , g_k and h_k	112
4.8.3	Optimization Method	115
4.9	Experimental Results	117
4.10	Discussion and Conclusion	120

Chapitre 5

Multiples simples courbes fermées pour appariement de modèle libre avec images de scène semitransparente -

La technique de déformation

Article: *Multiple simple closed curve for free form model matching with images from semitransparent scene*

		130
5.0	Résumé	130
5.1	Introduction	136
5.2	Problem statement	139
5.3	Related work	141
5.4	Method and assumptions	144
5.5	Object modeling	147
5.5.1	<i>Simple closed curve</i> - Foundations	148
5.5.2	Extremal and inner curves	149
5.6	Probabilistic model	150
5.6.1	Global criterion	151
5.6.2	Objective function with extremal and inner curves	154
5.6.3	Data driven term	155
5.6.4	Model driven term	157
5.6.4.1	What kind of solution-model consistency measurement?	158
5.6.4.2	Statistical representation	159
5.7	Search method	162
5.7.1	Deformation technique	165
5.7.1.1	Segment to deform	166
5.7.1.2	Generation of the new segment	167
5.7.2	Algorithmic constraints	169
5.7.3	Functional constraint - Criterion value variation	173
5.8	Implementation and experimental results	174
5.8.1	Implementation	174
5.8.2	Results	176
5.9	Discussion and conclusions	178

Chapitre 6	
Synthèse et discussion	194
6.1 Synthèse	194
6.2 Voies futures de recherche	200
6.2.1 Représentation reliée à l'observateur	202
6.2.2 La préservation du squelette	203
6.2.3 Mesure d'adéquation basée sur l'appariement point-à-point	204
6.2.4 La recherche par raisonnement évidentielle	205
6.3 Discussions et recommandations	206
Conclusion	210
Références	215
Annexes	229

Liste des tableaux

Tableau 6.1. Objectifs, difficultés et hypothèses.	195
Tableau 6.2. Notre approche.	196
Tableau 6.3. Les contraintes du processus de déformation progressive.	196
Tableau 6.4. Alignement global du modèle.	197
Tableau 6.5. Les contributions de la thèse.	199

Liste des figures

Figure 2.1. Exemple de projection radiographique postéro-antérieure de vertèbres humaines.	25
Figure 2.2. Résultat de la détection des discontinuités de la <i>figure 2.1</i> . Le passage par zéro de la convolution avec le Laplacien d'une Gaussienne a été masqué par le gradient binarisé à un seuil de 0.2.	26
Figure 2.3. Projection radiographique d'un objet isolé et détection des discontinuités utilisant les mêmes paramètres qu'à <i>figure 2.2</i>	26
Figure 2.4. Modèle <i>simple</i> de multiples SCF (MMC <i>simple</i>) à deux courbes extrémales (à gauche). MMC <i>complet</i> pour représentation de projections d'objets volumétriques à deux courbes extrémales et deux courbes internes (à droite).	34
Figure 3.1. Attraction neighborhood δs centered on site s of the transformed model $\xi(M, \alpha)$. The attraction force $\vec{F}_{s,t}$ between s and the boundary element $\vec{b}(Y_t)$ of the observed image Y at the site t is aligned with the unit vector $\vec{u}_{s,t}$	75
Figure 3.2. The dynamic model. $\vec{b}(Y_{\delta s, t})$ is 1-D slice of the observed local directional gradient. Left: surfaces 1 and 2 are equals, $\vec{F}_s = 0$ and the wire frame element s is in an equilibrium state over s' . Right: the equilibrium is disturbed and is shifted from s' ; surface 1 is greater than surface 2, hence left component \vec{F}_s of is greater than the right one, and s will be attracted to s'	75
Figure 3.3. To compute the stress \vec{S}_α in the direction \vec{u}_α , the model plane is split into two half planes P_α^+ and P_α^- centered on the gravity center of the model	75
Figure 3.4. In the upper panels, the wire frame models are superimposed on two noisy synthetic images. The lower panels show the final solutions after 200 time steps superimposed on the corresponding noisy images.	76
Figure 3.5. Evolution of the transformation parameters of the second alignment experiment of <i>Figure 3.1</i> . At the top left, top right, bottom left and bottom right we have the and translations, the rotation and the corresponding trend of the four terms of matrix G , respectively.	77
Figure 3.6. <i>a</i> and <i>b</i> : 2-D wire frame model superimposed over two X-ray images of a vertebra phantom taken from two slightly different poses. <i>c</i> and <i>d</i> : The corresponding alignment results after 200 time-steps. A 2-D articulated model is used for the result of <i>d</i>	78
Figure 3.7. Binary image alignment. <i>a</i> : club image. <i>b</i> : the affine transformation of <i>a</i> . <i>c</i> :	

alignment result of b to a after 200 time-steps. d : difference between a and c	79
Figure 3.8. a and b : two different views of a toy scene. c : the alignment result of b to a after 200 time-steps. d : difference between a and c	80
Figure 3.9. a and b : Lenna image and its affine transformation. c : the alignment result of b to a after 200 time-steps. d : difference between a and c	81
Fig. 4.1. Semitransparent 2-D projection of multiple contours. Each line l_k and its dual region r_k are defined on binary and discrete lattices L_k^l and L_k^r , respectively. The superpositions of the K curves and the K regions are denoted l and r defined on the K -ary lattices L^l and L^r , respectively. Here $K=3$	125
Fig. 4.2. The neighborhood system δs at site s of l_k is partitioned by segment $l_{k,\delta s}$ into subsystems $\delta s_{in(k)}$ and $\delta s_{out(k)}$. Labels α_1 and α_2 are defined on the region process $r_{\setminus k}$	126
Fig. 4.3. Crossing curves: $s \in l_k$, δs is the neighborhood of s , and l_h crosses l_k . α_i and α_j are labels in process $r_{\setminus k}$	126
Fig. 4.4. Two non-crossing lines l_k and l_h	126
Fig. 4.5. a : The adjacency graph of the 2-D model. A node represents a connected region and a branch links two adjacent regions. The node 0 indicates the background, and nodes 3' and 3'' denote two disjoint regions having the same overlapping combination. b and c : The region and boundary segmentation (12 boundary labels) of the model coded in grey level.	127
Fig. 4.6. The histogram distribution (dashed lines) of the local angular direction of the model's segments 1 to 4. The thresholded window type curves (solid lines) are generated from histograms. The abscissa represents the discretized local angular direction of 5° per step.	127
Fig. 4.7. $a0$: ST projection of three uniform objects over linearly variable background. $a1$ and $a2$: i.i.d. $N(\mu=0; \sigma=1.0 \text{ and } 3.0)$ noise added to $a0$. $b0$, $b1$ and $b2$: results of matching with $a0$, $a1$ and $a2$	127
Fig. 4.8. $a1$: The subobject frontiers are set to 1.0 and the inner regions are set to 0.2. $a1$ and $a2$: i.i.d. $N(\mu=0; \sigma=1.0 \text{ and } 1.5)$ noise added to $a0$. $b0$, $b1$ and $b2$: results of matching with $a0$, $a1$ and $a2$	128
Fig. 4.9. $a1$ and $a2$ are generated from $a0$ by adding an i.i.d. $N(\mu=0; \sigma)$ noise of $\sigma=0.8$, and 1.2, respectively. $b0$, $b1$ and $b2$: results of matching with $a0$, $a1$ and $a2$	129
Fig. 4.10. Using the same parameters, model and start state, as before except β_k is set to zero, $b1$ and $b2$ are the results of matching with $a1$ of Fig. 4.8 and $a2$ of Fig. 4.9, respectively.	129
Fig. 5.1. Cylindric, conic and planar free forms. The 2-D semitransparent projection of these forms from a general viewing position can be described using one extremal curve with one or two inner curves.	183
Fig. 5.2. Topologic relations between extremal curve l_k and inner curve l_{k1} . l_{k1}	

transforms the domain r_k from a *simply* connected region to *doubly* connected (left). Here, l_{k1} is considered as a hole in r_k . In the right scheme, $l_k \cap l_{k1} = \emptyset$. Hence, l_k and l_{k1} can be partitioned into one *simple* segment and one *double* segment, as

$$l_k = \{l_k^{\setminus k1}, l_k^{k1}\} \text{ and } l_{k1} = \{l_{k1}^{\setminus k}, l_{k1}^k\}. \quad 183$$

Fig. 5.3. Partition of the local neighborhood system δs for extremal-inner curves configurations. l_k , l_{k1} and l_h are *extremal*, *inner*, and additional curves, respectively. v_s is given by the union of the shaded parts of δs . When l_h does not cross l_k or l_{k1} (a2, b2 and c2) the sum of the weighting coefficients $\mu_s = \sum_{j=1, \dots, m} \mu_{s,j}$ equals to v_s , minus the surface of the + shaded part. 183

Fig. 5.4. Curve segmentation and labeling of the curve set $l = \{l_0, l_1, l_2\}$. Eight labeled segments are detected $l = \{l_0^{(b_1)}, l_0^{(b_2)}, l_0^{(b_3)}, l_0^{(b_4)}, l_1^{(b_5)}, l_1^{(b_6)}, l_2^{(b_7)}, l_2^{(b_8)}\}$ 184

Fig. 5.5. Typical situation where the convergence is compromised by the *gluing* drawback the model driven term. The model and the intensity image are shown in solid lines and shaded regions, respectively. The curve l_h can not match exactly the real

discontinuities because of the uninodality of $H_0(l_k^{M(b_2)}, \theta)$ 185

Fig. 5.6. The global search method. 185

Fig. 5.7. Deformation technique with two sampling processes for s and s_a . The summit of the deformation, s_a , is sampled from uniform or intensity gradient distribution. s_a is confined to the maximum deformation disk $\Delta_{s,d}$ 186

Fig. 5.8. Determination of the curve segment to deform $l_{k,(s_0,s_1)}$. The curve l_k is scanned *forward* and *backward* from s . $l_{k,(s_0,s_1)}$ is confined to the disk $\Delta_{s_a,d}$. The summit s_a can belong either to the outer region (left) or to the inner region (right) of l_k 186

Fig. 5.9. Detection of the segment to deform $l_{h,(t_0,t_1)}$ of the *inner* curve l_h when the segment to deform $l_{k,(s_0,s_1)}$ of the *extremal* curve l_k contains a *splitting* point s_{SM} . 187

Fig. 5.10. Generation of the new segment $l_{k,(s_0,s_a,s_1)}$ using two *glued* segments of cubic spine $l_{k,(s_0,s_a)}$ and $l_{k,(s_a,s_1)}$ 187

Fig. 5.11. Generation of the new *doubly* connected pair $l_{k,(s_0,s_a,s_1)}$ and $l_{h,(t_0,t_{aSM},t_1)}$. t_{SM} is the old *splitting* point and t_{aSM} is the generated *splitting* point. 188

Fig. 5.12. Three schemes where the curve deformation $\delta l_k = l_{k,(s_0,s_1)} - l_{k,(s_0,s_1)}$ locally fulfill conditions 1 and 2. $l_{k,(s_0,s_1)}$: dashed line; $l_{k,(s_0,s_1)}$: solid line. Left: $r_{\setminus k}^{(a_{k,i})}$ is a *simply* connected region. Center and right: $r_{\setminus k}^{(a_{k,i})}$ is a *doubly* connected region enclosing a hole of one connected region (one simple closed curve) and a hole of three connected

regions (two simple closed curves), respectively.	188
Fig. 5.13. Free conic (left) and free cylindric (right) object used to match with noisy synthetic images.	190
Fig. 5.14. Experiments with synthetic images modeling semitransparent projections of an arbitrary free conic object. Images a_1 and a_2 are generated by adding an i.i.d. $N(\mu=0; \sigma)$ noise of $\sigma=0.5$ and 1.0 to a_0 , respectively. b_0 , b_1 and b_2 are the corresponding matching results with a_0 , a_1 and a_2 , respectively.	190
Fig. 5.15. Experiments with synthetic images modeling semitransparent projections of an arbitrary free cylindric form. Images and results are arranged in the same order of Fig. 5.14.	190
Fig. 5.16. Two hand-drawn wire-frame models to match with X-ray images.	191
Fig. 5.17. Experiments with three X-ray images of the a synthetic object. Images a_0 , a_1 and a_2 have the same 2-D relational graph. b_0 , b_1 and b_2 are the matching results of a_0 , a_1 and a_2 with model a of Fig.5.16 after 5 full sweep iterations, respectively.	191
Fig. 5.18. Experiments with three X-ray images of the same object. Images a_0 , a_1 and a_2 have the same 2-D relational graph. Using the hand-drawn model b of Fig. 5.16, b_0 , b_1 and b_2 are the corresponding matching results after 5 full sweep iterations, with a_0 , a_1 and a_2 , respectively.	192
Fig. 5.19. Experiments with noise-corrupted images (see the text). b_0 , b_1 and b_2 are the corresponding matching results with a_0 , a_1 and a_2 , respectively.	193
Figure 6.1. Graphes 2-D de deux projections issues à partir de deux directions voisines.	202
Figure 6.2. L'intégration d'une mesure de similarité de squelette dans une structure hiérarchique réduit la probabilité de déstructuration de la solution.	203
Figure B.1. Les règles de Gestalt pour l'organisation perceptuelle. Soulignons que les règles n'ont pas toutes la même importance. En cas de conflit, le regroupement se fait suivant la règle la plus dominante. Dans l'expérience de la symétrie, vérifiez que le regroupement se fait entre des lignes symétriques et non entre des lignes voisines.	233

Liste des symboles

Chapitre 3 (article 1)

Y	image observée
X	image modèle à aligner
M	modèle à aligner
C^M	modèle de multiples simples courbes fermées
K	nombre de courbes
a	vecteur des paramètres de transformation
$\xi(M, a)$	fonction de transformation du modèle M utilisant les paramètres a
ϕ et ψ	fonctions d'interactions locales entre Y et $\xi(M, a)$
s et t	sites de la grille de discrétisation
δs	système de voisinage centré sur s
$\vec{b}()$	vecteur de gradient directionnel
$V_{s,t}$	mesure de similarité entre deux sites
$\vec{F}_{s,t}$	force d'attraction entre s et t
\vec{F}_s	force locale appliquée sur s
F	champ de force distribué
$\vec{u}_{s,t}$	vecteur unitaire de s à t
\vec{u}_α	vecteur unitaire dans la direction α
s_{GC}	centre de gravité
m	masse
J	moment d'inertie
f	coefficient de dissipation ou de frottement
l	vecteur de translation
$\bar{\theta}$	angle de rotation autour du centre de gravité
α	direction de mise à l'échelle
\vec{g}_α	vecteur de mise à l'échelle dans la direction α
\vec{F} et \vec{F}^N	force résultante de translation et force seuillée
\vec{C} et \vec{C}^N	couple résultant de rotation et couple seuillé

\bar{S}_α et \bar{S}_α^N	tension dans la direction α et tension seuillée
R_α	transformation de rotation d'angle α
$G_{\alpha,g}$	transformation de mise à l'échelle de valeur g dans la direction α
G	transformation résultante d'une série de mises à l'échelle
P_α^+ et P_α^-	demi-plan positif et demi-plan négatif définis par la direction α
t	variable du temps
N_S	nombre des directions sur lesquelles la tension est calculée
th_F et th_C	seuil de la force de translation et seuil du couple de rotation

Chapitre 4 (article 2)

l	processus aléatoire des lignes
l_k	processus aléatoire de la k ème ligne
r	processus aléatoire des régions
r_k	processus aléatoire de la k ème région
l	réalisation du processus des lignes ou l'état courant de la solution
l_k	réalisation du k ème processus de ligne ou k ème ligne
r	réalisation du processus de région ou l'état présent de la solution
r_k	réalisation du k ème processus de région
$in(l_k)=in(r_k)=in(k)$	région interne de la k ème ligne
$out(l_k)=out(r_k)=out(k)$	région externe de la k ème ligne
K	nombre de contours dynamiques formant le modèle
Y	image observée
L	grille de discrétisation
$c= L $	nombre des sites dans L
s et t	sites de L
δs	système de voisinage centrée sur s
$\delta s \setminus s$	système de voisinage centrée sur s dépourvu du site s
$\delta s_{in(k)}$ et $\delta s_{out(k)}$	sous-voisinages de δs définis sur la région interne et la région externe de l_k
$l_{\cdot,s}$ et $r_{\cdot,s}$	état courant de l'ensemble des lignes et l'ensemble des régions au site s
$l_{k,s}$ et $r_{k,s}$	k ème ligne et k ème région au site s
$l_{\setminus k}$ et $r_{\setminus k}$	ensemble des lignes (régions) sans la ligne l_k (région r_k)
$a=\{a_1, \dots, a_m\}$	ensemble d'étiquettes locales définies dans $r_{\setminus k, \delta s}$
a_s	étiquette dans $r_{\setminus k, \delta s}$ au site s
$r_{\setminus k, \delta s}^{(a_j)}$ et $\delta s^{(a_j)}$	partie de $r_{\setminus k, \delta s}$ et sous-voisinage de δs ayant l'étiquette a_j dans $r_{\setminus k, \delta s}$

ρ, ρ_s et $\rho_{s,j}$	coefficients de pondérations reliés à $\delta s, \delta s^{(a_s)}$ et $\delta s^{(a_j)}$
$\bar{Y}_{\delta s_{in(k)}^{(a_j)}}$ et $\bar{Y}_{\delta s_{out(k)}^{(a_j)}}$	moyennes de l'image observée calculées sur $\delta s_{in(k)}^{(a_j)}$ et $\delta s_{out(k)}^{(a_j)}$
$\theta_{k,s}$	angle de la direction locale à $l_{k,s}$
$d^n \theta_{k,s}$	dérivée curviligne d'ordre n de $\theta_{k,s}$
$v_{k,s} = \{\theta_{k,s}, d\theta_{k,s}, d^2\theta_{k,s}, \dots, d^p\theta_{k,s}\}$	descripteurs locaux de $l_{k,s}$
E	fonction d'objectif de trois termes
f_k, g_k et h_k	terme relié à l'image, terme relié au modèle et terme de lissage de l_k
$\alpha_{k,s}, \beta_{k,s}$ et $\gamma_{k,s}$	coefficients de pondération de f_k, g_k et h_k
$\gamma_{k,s,1}, \gamma_{k,s,2}$ et $\gamma_{k,s,3}$	coefficients de pondération des composantes du terme de lissage
$O = \{O_1, \dots, O_K\}$	objet formé de la superposition semitransparente de K sous-objets
G_O	graphe relationnel décrivant les relations de voisinage des régions de O
R_O et n_R	ensemble des noeuds de G_O et nombre des noeuds de G_O
F_O et n_F	ensemble des branches de G_O et nombre des noeuds de G_O
$\xi()$	fonction de formation d'image à partir du modèle sous-jacent
$N(\mu, \sigma)$	bruit additif, indépendant et identiquement distribué de moyenne μ et de variance σ^2
$\Delta_{s,d}$	disque centré sur le site s de rayon d
C_k	contour dynamique
$C_{k,\Delta_{s,d}} = C_k \cap \Delta_{s,d}$	segment de contour résultant de l'intersection de C_k avec $\Delta_{s,d}$

Chapitre 5 (article 3)

La notation des symboles dans l'article 3 est conforme à celle de l'article 2. Cependant, certains symboles additionnels sont utilisés uniquement dans l'article 3 et sont décrits ci-après.

l_k et l_h	courbes extrémales
l_{k1} et l_{k2}	courbes internes
$r_k^I = in(r_k)$ et $r_k^O = out(r_k)$	intérieur et extérieur de la région r_k
$\bar{in}(l_k) = in(l_k) \cup l_k$	ensemble fermé de la courbe l_k et de son intérieur
l_k^{k1}	segment de courbe extrémale l_k qui est commun à la courbe interne l_{k1} : segment <i>double</i>
$l_k^{\setminus k1} = l_k \setminus l_k^{k1}$	segment de courbe extrémale l_k qui n'est pas commun à la courbe interne l_{k1} : segment <i>simple</i>
l_{k1}^k	segment de courbe interne l_{k1} qui est commun à la courbe extrémale l_k :

	segment <i>double</i>
$l_{k1}^{\setminus k} = l_{k1} \setminus l_{k1}^k$	segment de courbe interne l_{k1} qui n'est pas commun à la courbe extrême l_k : segment <i>simple</i>
$l_{k,(s_0,s_1)}$	segment de courbe l_k partant du site s_0 et finissant au site s_1 dans le sens de l'aiguille de la montre
l^M	ensemble de multiples courbes formant le modèle
l^0	ensemble de multiples courbes formant la solution initiale
$E(l, Y, l^M)$	fonction d'objectif
$E_D(l, Y), E_M(l, l^M)$ et $E_S(l)$	terme relié à l'image observée, terme relié au modèle et terme de lissage
δE	variation de valeur de la fonction d'objectif due à une déformation de courbe
$l_{\setminus k \setminus k1}$ et $r_{\setminus k \setminus k1}$	ensemble des lignes (régions) sans les lignes l_k et l_{k1} (régions r_k et r_{k1})
$A(r_{\setminus k, \delta s}) = a = \{a_1, \dots, a_m\}$	opération d'extraction d'étiquettes de la région $r_{\setminus k, \delta s}$
$A_{\setminus k}(l_{k,(s_0,s_1)}) = \{a_{k,1}, a_{k,2}, \dots, a_{k,n}\}$	opération d'extraction d'étiquettes des régions de $r_{\setminus k}$ transversées séquentiellement par $l_{k,(s_0,s_1)}$
$S(l_k)$	$= l_k$ lorsque l_k est une courbe extrême et $= l_{k1}^{\setminus k}$ lorsque l_{k1} est une courbe interne de l_k
a_j	étiquette d'une région ou d'un noeud du graphe G
b_i	étiquette d'un segment de courbe ou d'une branche du graphe G
$A^M = \{a_j, j=1, \dots, J\}$	ensemble des noeuds du graphe relationnel G^M de l'objet M
$B^M = \{b_i, i=1, \dots, I\}$	ensemble des branches du graphe relationnel G^M de l'objet M
$l_k^{(b_i)}$	segment de courbe l_k d'étiquette b_i
$H_n(l_k^{(b_i)}, d^n \theta)$	histogramme de la mesure locale de $d^n \theta$ sur $l_k^{(b_i)}$
$H_n^{G_o}(l_k^{(b_i)}, d^n \theta)$	$= H_n(l_k^{(b_i)}, d^n \theta)$ filtré par un gaussien G_o
$WH_n^{G_o}(l_k^{(b_i)}, d^n \theta)$	$= H_n^{G_o}(l_k^{(b_i)}, d^n \theta)$ seuillé
$\ l_{k,(s_0,s_1)}\ $	longueur curviligne du segment de courbe $l_{k,(s_0,s_1)}$
s_a	sommet de déformation
s_{SM}	site de jonction entre une courbe extrême et une de ses courbes internes
s_{aSM}	nouveau site de jonction entre une courbe extrême et une de ses courbes internes
$\Delta_{s_a, d}$	disque centré sur le site s_a de rayon d

$l_{k,\Delta_{s_a,d}}$	intersection de la courbe l_k avec le disque $\Delta_{s_a,d}$
$l_{k,(s_0,s_a,s_1)}$	nouveau segment de courbe l_k commençant par le site s_0 , finissant par le site s_1 et passant par le sommet de déformation s_a
$\delta l_k = l_{k,(s_0,s_1)} - l_{k,(s_0,\dots,s_1)}$	déformation du segment de courbe $l_{k,(s_0,s_1)}$ à $l_{k,(s_0,\dots,s_1)}$
δr_k	changement de région produite par la déformation δl_k
$G_{\setminus k}$	graphe relationnel de région $r_{\setminus k}$
$sG_{\setminus k} \subset G_{\setminus k}$	sous-graphe isolé du graphe $G_{\setminus k}$ et lié au restant de $G_{\setminus k}$ à travers un seul noeud
$AS1(\delta l_k) = l_{k,(s_0,\dots,s_1)} \cup l_{k,(s_0,s_1)}$, $AS2(\delta l_k) = l \cap \delta r_k$ et $AS3(\delta l_k) = \{s \mid \exists t \in \delta s ; t \in \delta r_k\}$	trois régions affectées par la déformation δl_k et utilisées pour calculer δE

Liste des annexes

Annexe A	
Stabilisateurs fonctionnels	229
Annexe B	
Contraintes géométriques	232
Annexe C	
Modèles élastiques	235

Préface

La scoliose idiopathique est une maladie qui génère des déformations tridimensionnelles (3-D) complexes de la colonne vertébrale et de la cage thoracique et dont l'étiologie demeure méconnue. La reconstruction 3-D de la colonne vertébrale et de la cage thoracique est essentielle, entre autres, au diagnostic précis de la maladie, à l'évaluation des traitements et au suivi de l'évolution de la maladie. L'utilisation des méthodes de reconstruction par coupes transversales, comme celles issues de la tomographie axiale, était écartée principalement à cause du besoin d'effectuer la reconstruction 3-D des structures osseuses en position debout, et aussi afin d'éviter des fortes doses de rayonnements ionisants (Barrett et Swindell, 1981). Notre groupe de recherche a développé une technique de reconstruction basée sur l'inférence par stéréo-correspondance à partir d'un nombre limité de projections radiographiques issues de différentes directions (de Guise *et al.*, 1995; Dansereau et Stokes, 1988; Marzan, 1976). Cependant, cette technique présente plusieurs limitations: (i) elle est basée sur l'appariement d'un nombre limité de repères anatomiques; (ii) la détection et l'appariement de ces repères sont faits manuellement; (iii) la reconstruction se fait par positionnement et ajustement global de modèle géométriques de vertèbres aux points 3-D inférés par l'appariement des repères détectés.

On a identifié un besoin de développer un système de vision artificielle pour la détection et la reconnaissance d'objet dans les images radiographiques. Idéalement, ce système devrait

détecter et identifier avec précision et robustesse les différentes structures anatomiques nécessaires à la reconstruction et à la représentation 3-D de la colonne vertébrale, de la cage thoracique et du bassin. Le travail proposé est un premier pas dans cette direction.

La thèse est rédigée sous forme d'articles et est organisée de la façon suivante:

1. le premier chapitre permet d'introduire la problématique de la vision artificielle et de l'imagerie médicale et de situer le travail dans le contexte général du traitement des images radiographiques du rachis humain;
2. le second chapitre expose l'approche globale basée sur un modèle probabiliste de simples courbes fermées et déformables. Il permet aussi de situer et de résumer les trois articles formant le corps de la thèse et qui sont présentés successivement aux chapitres 3, 4 et 5.
3. l'article 1 (chapitre 3) présente une nouvelle technique d'alignement global d'un modèle de courbe dans une image observée. Cette technique servira à produire une estimation initiale de la solution nécessaire à la suite de la méthode. Les articles suivants sont consacrés à la formulation du modèle probabiliste des simples courbes fermées;
4. l'approche générale est développée dans l'article 2 (chapitre 4) pour un modèle simple formé uniquement de courbes extrémales. Cette approche est validée dans le contexte d'images synthétiques fortement bruitées;
5. le troisième article (chapitre 5) présente un modèle complet permettant le traitement d'images issues de la projection semitransparente d'objets volumétriques connus *a priori*. L'intérêt de cette méthode est aussi illustré par son application à des images synthétiques et à des images radiographiques d'objets connus simulant la géométrie des vertèbres humaines;

6. le chapitre 6 permet d'effectuer une synthèse des différents résultats obtenus et de discuter les éléments importants et originaux qui découlent de ce travail. On y retrouvera aussi une mise en perspective de la méthodologie développée, une série de questions soulevées par ce travail et une série de recommandations pour la poursuite de celui-ci.

Enfin, une conclusion permet de rappeler les principaux objectifs, les grandes lignes de la méthodologie et les contributions majeures de ce travail. Le lecteur pourra aussi retrouver en annexes des informations techniques et méthodologiques complémentaires permettant la lecture éclairée de cette thèse.

Chapitre 1

Introduction

Depuis très longtemps, les philosophes et les scientifiques ont été fascinés par la faculté humaine reliée à la perception et à la compréhension des objets qui nous entourent. La vision artificielle (ou vision par ordinateur) est une première tentative de doter la machine, en l'occurrence l'ordinateur, de certaines des propriétés de l'intelligence humaine en ce qui concerne les capacités d'analyser, de reconnaître et de décrire les différentes caractéristiques d'une scène tridimensionnelle à partir d'un ensemble limité d'images de cette scène. Les travaux présentés dans cette thèse relèvent en partie de cette problématique d'interprétation et de reconnaissance d'objets à partir d'images représentant de façon incomplète et imparfaite une scène tridimensionnelle. Une autre préoccupation fort ancienne de l'activité scientifique est la compréhension du fonctionnement du corps humain et la connaissance complète et intime de chacune de ses structures et de ses organes. L'imagerie médicale essaie d'apporter une réponse partielle à ce questionnement en fournissant des moyens de sonder l'intérieur du corps humain de façon de moins en moins invasive et de plus en plus précise et spécifique. Les trente dernières années ont vu l'apparition et l'utilisation à grande échelle des techniques de radiologie et de tomographie axiale par ordinateur (CAT SCAN) qui utilisent les principes du rayonnement des particules et les principes de la résonance magnétique en leur associant des

méthodes informatiques. Les travaux présentés dans cette thèse relèvent aussi de l'utilisation des méthodes d'imagerie médicale, en particulier de la radiographie de projection, dans le but de reconnaître et de reconstruire en 3-D les organes du corps humain.

Les applications de la vision par ordinateur dans le domaine de l'imagerie médicale sont de plus en plus nombreuses. On peut les regrouper en deux ensembles selon qu'elles fassent appel au niveau primaire ou au niveau intermédiaire de la vision artificielle. Dans le premier groupe, on retrouve les applications utilisant les techniques de traitement local de l'information centrées soit sur le pixel, soit sur une région délimitée de l'image. Dans le second groupe, les applications font appel à des informations de niveau plus global et structurel, organisées selon des modèles descriptifs ou topologiques connus *a priori*. Parmi les applications du premier groupe, on retrouve les techniques basées sur l'extraction des régions homogènes (seuillage de l'histogramme des intensités de gris dans l'image), la détection des discontinuités d'intensité dans l'image basée sur les informations de gradient (Kass *et al.*, 1988; Leymarie et Levine, 1993; Ayache *et al.*, 1992) ou l'analyse de mouvement et la caractérisation de tissus à partir de séquences d'images (Nastar et Ayache, 1996; Bardinet *et al.*, 1996). Parmi les applications du deuxième groupe, on retrouve les méthodes basées sur la reconnaissance d'objets, l'étiquetage des régions (Suetens *et al.*, 1988), l'appariement avec un modèle (atlas anatomique ou modèle fonctionnel) (Bajscy et Kovacic, 1989) et la reconstruction 3-D (Close, 1991; Hoffman *et al.*, 1988; Dansereau *et al.*, 1990).

Dans cette thèse, nous développons un système de traitement de niveau intermédiaire pour la détection et l'étiquetage d'images de scènes semitransparentes (SST). Rappelons que

l'ensemble de notre travail s'intègre dans une approche globale ayant pour but la reconstruction 3-D des structures osseuses de la colonne vertébrale à partir d'un nombre limité de vues radiographiques (de Guise *et al.*, 1995; Dansereau *et al.*, 1990; Dansereau et Stokes, 1988; Marzan, 1976). Les résultats fournis par notre approche de détection et d'étiquetage serviront pour l'étape de reconstruction 3-D (Rosenfeld, 1988; Dansereau *et al.*, 1990; Aubin, 1995). Dans ce cas, la reconstruction se fera par positionnement et déformation de modèle (Trochu, 1993; Dansereau *et al.*, 1990).

Ce chapitre est organisé en quatre sections. Dans la première section, nous abordons la problématique générale de la vision artificielle en introduisant l'aspect mal posé du problème inverse. Nous présentons le processus de reconnaissance d'objet comme un problème de recherche dans un espace d'appariement entre le modèle et l'image. Puis nous discutons des contraintes organisationnelles qui sont généralement utilisées pour changer la nature sous-contrainte du problème et pour réduire les dimensions de l'espace de recherche. La deuxième section sera consacrée à la présentation de la problématique particulière d'un système de vision artificielle appliquée à des images de scène semitransparentes dans le contexte de l'imagerie médicale. Dans la troisième section, nous définissons les objectifs et les hypothèses générales de notre travail. Enfin, la quatrième section présente de façon synthétique ce qui constitue le corps principal de la thèse.

1.1 Problématique générale

Puisque l'objectif général de ce travail est de développer un système de vision artificielle

appliqué aux images radiographiques, nous présentons d'abord les objectifs globaux d'un système de vision artificielle ainsi que la structure hiérarchique généralement utilisée dans plusieurs modèles de vision. Le problème de la vision artificielle étant un problème mal posé (au sens mathématique du terme), les méthodes de régularisation par stabilisateurs fonctionnels seront aussi introduites. Les notions d'espace de recherche, de convexité et de dimension de cet espace seront aussi abordées en fonction des différentes méthodes de recherche. Nous discutons aussi de l'utilisation de contraintes naturelles de l'organisation perceptuelle dans le but de la réduction de l'espace de recherche.

1.1.1 Vision artificielle

Le problème de la vision artificielle peut être présenté de façon analogue au problème inverse de l'optique géométrique (Marroquin *et al.*, 1987). Le problème de l'optique géométrique classique (ou du graphisme par ordinateur) est la construction d'une image 2-D par le calcul de la projection de la scène 3-D sur le plan de l'image en tenant compte de la géométrie de la scène, du type de la projection (orthographique ou perspective) et de la pose (point de vue). L'objectif principal de la vision artificielle est la restitution d'une scène 3-D par l'analyse et l'interprétation d'une ou de plusieurs images 2-D de cette scène (Poggio *et al.*, 1985). La vision artificielle comprend aussi l'étude et l'implantation des différents processus permettant l'automatisation de l'interprétation des images (Marr, 1982).

Plusieurs systèmes de vision artificielle proposent une structure hiérarchique à trois niveaux pour le traitement de l'information provenant des images: un niveau inférieur ou primaire, un niveau intermédiaire et un niveau supérieur. Le niveau primaire regroupe

habituellement les opérations portant sur l'extraction des primitives géométriques simples comme la détection des discontinuités et l'isolation des régions homogènes. Les processus d'analyse de niveaux plus élevés utilisent des connaissances et des regroupements perceptuels plus structurels (Marr, 1982). Le niveau intermédiaire peut contenir la description et l'étiquetage des différentes régions segmentées et la reconstruction des surfaces et des objets de la scène. Enfin, le niveau supérieur fournit habituellement une description et une interprétation complète de la scène incluant la reconstruction géométrique complète en 3-D. Toutefois, les limites entre les différents niveaux ne sont pas bien définies et il est parfois difficile de bien identifier le niveau auquel intervient la reconnaissance de l'objet, soit avant (niveau primaire) ou après (niveau supérieur) la reconstruction 3-D (Rosenfeld, 1988). En effet, dans la vision humaine, le raisonnement et la résolution des problèmes sont présents à tous les niveaux de l'analyse perceptuelle (Rock, 1975; 1983).

La première étape du traitement des informations visuelles passe généralement par l'extraction des primitives en vue d'obtenir une description compacte de l'image. De plus, cette description doit idéalement être complète et significative (Torre et Poggio, 1986). Elle sera complète si toute l'information contenue dans l'image est représentée. Elle sera significative si elle reflète l'ensemble des propriétés de la scène 3-D.

La méthode classique permettant d'obtenir une première description compacte de l'image utilise la détection des discontinuités d'intensité de l'image. La détection de discontinuités consiste à identifier les changements d'intensité dans l'image suivant les processus physiques qui les ont engendrés (Torre et Poggio, 1986). Ce processus de détection permet de simplifier l'analyse d'images en réduisant radicalement la quantité d'information à traiter tout en

préservant l'information relative aux limites et aux frontières des objets constituant la scène (Canny, 1986). L'importance de ce processus est justifiée par la propriété d'invariance des discontinuités par rapport aux changements dans les paramètres de formation d'image comme, par exemple, le point de vue et les conditions d'éclairage.

Un détecteur de discontinuités doit idéalement satisfaire à trois critères de base:

1. une sensibilité élevée; le détecteur doit révéler presque toutes les arêtes de discontinuités dans l'image;
2. une grande précision pour la localisation des arêtes détectées; cette précision est critique pour la mesure des disparités utilisées habituellement par les méthodes d'inférence des formes à partir du mouvement ou de la stéréo correspondance;
3. une robustesse de détection par rapport au bruit; ce critère permet d'assurer la stabilité des étapes subséquentes d'analyse.

Un tel détecteur de discontinuité permet alors d'extraire une description compacte, complète et significative de l'image (Torre et Poggio, 1986).

1.1.2 Le problème inverse

Le processus de projection d'une scène 3-D dans une image discrète 2-D cause la perte d'une importante quantité d'information (Marroquin *et al.*, 1987). Du point de vue géométrique, la fonction de projection réduit la dimension de l'espace de trois à deux. Par conséquent, la reconstruction de la réalité 3-D uniquement à partir de quelques projections conduit généralement à des solutions ambiguës. Ici, on parle d'un problème inverse sous-contraint (Bertero *et al.*, 1988). Afin de lever les ambiguïtés, il faut introduire des contraintes

additionnelles et naturelles basées sur des propriétés connues du monde physique. Cela est typique à plusieurs problèmes inverses en mathématique et en physique. En effet, la caractéristique commune à la plupart des problèmes de la vision primaire peut être formulée de la façon suivante (Poggio *et al.*, 1985; Bertero *et al.*, 1988): le problème de la vision primaire est mal posé selon le sens défini par Hadamard. Un problème est dit bien posé si sa solution (i) existe, (ii) est unique, et (iii) dépend continûment des données observées. Les problèmes mal posés ne répondent pas à un ou plusieurs de ces critères.

Résoudre un problème inverse mal posé passe principalement par la réduction du nombre de solutions possibles en introduisant convenablement des connaissances *a priori* supplémentaires. Ces connaissances se traduisent généralement par les contraintes additionnelles ajoutées au problème pour assurer l'existence, l'unicité et la stabilité des solutions (Terzopoulos, 1986). L'utilisation des méthodes de régularisation est une façon de stabiliser le problème en imposant des contraintes fonctionnelles de lissage. L'annexe A présente un aperçu des stabilisateurs fonctionnels.

Les méthodes usuelles de régularisation des problèmes mal posés (Tikhonov et Arsenin, 1977) proposent de trouver la solution y à partir des observations Y sachant que $A.y = Y$. Ici A est un opérateur linéaire. En utilisant le stabilisateur quadratique $\|P.y\|^2$, la solution régularisée au sens de Tikhonov est donnée par la minimisation de la fonction suivante:

$$\Phi(y, Y, \lambda) = \|A.y - Y\|^2 + \lambda \|P.y\|^2 \quad (1.1)$$

λ est appelé le paramètre de régularisation. Dans cette méthode, λ ajuste le compromis entre le degré de régularisation de la solution et sa conformité aux données observées Y . P représente les contraintes physiques imposées au problème. Le premier terme de $\Phi(y, Y, \lambda)$

mesure la distance entre la solution et les données, alors que le deuxième terme représente une certaine fonction d'énergie de la solution y . Pour la fonction quadratique $\Phi(y, Y, \lambda)$, l'espace de solution est convexe et une solution unique existe (Marroquin *et al.*, 1987). Les méthodes de recherche peuvent être directes ou itératives. Les méthodes directes comme, par exemple, les algorithmes de convolution, sont généralement utilisées lorsque l'image est définie sur une grille régulière, et lorsque A (dans $\Phi(y, Y=A.y, \lambda)$) représente une opération invariante dans l'espace. Dans ce cas, la solution régularisée est obtenue en convoluant les données observées Y par le filtre inverse de A . Les méthodes itératives utilisées pour les problèmes convexes sont habituellement descendantes comme, par exemple, le gradient descendant ou le gradient conjugué. Elles peuvent être incorporées dans des structures parallèles et à multirésolution dans le but d'accélérer la convergence (Terzopoulos, 1986b).

Les méthodes usuelles de régularisation donnent des résultats satisfaisants lorsque la solution ne présente pas de grandes discontinuités. Cela est dû à la propriété de lissage des stabilisateurs quadratiques (Terzopoulos, 1986). Cependant, l'efficacité de ces méthodes en présence des discontinuités peut être améliorée si elles sont intégrées dans une structure hybride à trois étapes (Marroquin *et al.*, 1987): estimer la solution y en utilisant une régularisation usuelle; identifier les régions où l'erreur entre y et Y est élevée; reprendre l'estimation de la solution avec régularisation en utilisant les discontinuités pour déterminer les domaines d'intégration des stabilisateurs. Les fonctions de stabilisation non-linéaires présentées à l'annexe A peuvent aussi être utilisées pour tenir compte implicitement des discontinuités (Poggio *et al.*, 1985). Dans ce cas-ci, l'espace de solution n'est pas convexe

et possède plusieurs minima locaux. La méthode de recherche choisie doit avoir la capacité d'éviter les minima locaux.

1.1.3 L'espace de recherche

La reconnaissance de forme est un processus de niveau supérieur qui fait appel à une opération d'association de preuves entre l'image observée et un modèle choisi d'un ensemble d'objets possibles (Ballard et Brown, 1982). Ce processus se fait souvent en deux étapes (Ayache et Faugeras, 1986; Lowe, 1987, Brooks, 1981, 1986). La première étape accumule des preuves afin de proposer une hypothèse sur l'identité de l'objet à reconnaître et sur le modèle à utiliser. La deuxième étape vérifie l'hypothèse en appariant le modèle proposé avec les images observées. La décision de retenir ou de remplacer l'hypothèse dépend du résultat de l'appariement.

L'appariement se présente généralement comme un problème de recherche dans l'espace, noté S , des correspondances possibles entre les éléments de primitives du modèle et ceux extraits de l'image. La recherche d'une solution optimale dans S fait face généralement à deux principales difficultés: le temps de recherche et le choix de la méthode de recherche.

Le temps de recherche est une fonction monotone de la dimension de S . On suppose que le modèle M est représenté par n_M éléments de primitives notés $e_M = \{e_{M,1}, \dots, e_{M,n_M}\}$ et que les n_Y primitives $e_Y = \{e_{Y,1}, \dots, e_{Y,n_Y}\}$ extraites de l'image Y sont du même type que celles du modèle. L'espace de recherche $S(e_M, e_Y)$ n'est autre que l'ensemble de toutes les

combinaisons possibles $S(e_M, e_Y) = \{(e_{M,i}, e_{Y,j}); i=1, \dots, n_M; j=1, \dots, n_Y\}$ des appariements des éléments de e_M avec les éléments de e_Y . D'une manière générale, le cardinal de S est donné

par $|S| = \frac{n_Y!}{n_M!(n_Y - n_M)}$ si $n_M < n_Y$. Ainsi, $|S|$ est une fonction combinatoire du nombre des

primitives et est directement liée à la complexité de l'objet à reconnaître et à la quantité des primitives extraites de l'image. Il faut donc réduire n_M et n_Y . Quant au modèle, il s'agit de trouver une description représentative et compacte des objets à reconnaître. Cela nous amène aux techniques de reconnaissance évidentielle (ou par association de preuves) qui peuvent servir à développer des modèles de représentation hiérarchique utilisés dans des méthodes de recherche basées sur la classification par ordre de *confiance* (Chou et Brown, 1990; Lowe, 1985). L'étude de ces techniques ne fait pas partie de la perspective de cette thèse.

Du côté de l'image, le nombre des primitives extraites n_Y dépend de la nature de la scène et du processus de formation d'image. Dans le cas d'images bruitées et complexes, l'opération d'extraction de primitives, comme par exemple la détection des discontinuités d'intensité, fournira un ensemble d'éléments fragmentés. Cela fait augmenter considérablement le nombre n_Y . L'utilisation de l'organisation perceptuelle dans le regroupement des primitives permet de réduire n_Y et par conséquent l'espace de recherche.

Ce point sera discuté dans la section 1.1.4.

Quant au choix de la méthode de recherche, il doit se faire dans la perspective d'assurer la convergence vers une *solution optimale* et de réduire le *temps de recherche*. Le choix

dépend de la convexité¹ de l'espace de recherche S . La convexité de S est reliée directement à la fonction de transition définie par la méthode de recherche. La fonction de transition régit directement le système de voisinage défini dans l'espace S . Ainsi, deux états x et y de S sont dits voisins s'il existe un passage (une transition permise par la fonction de transition) entre les deux états. Un état x est dit un optimum local si pour tout état y plus probable que x , y ne peut pas être *rejoint* à partir de x par une série monotone de transitions (Hajek, 1988). Par conséquent, une fonction de transition riche en capacité de transition entre les états donne lieu à un système de voisinage dans l'espace S qui est large et dense. Ainsi, certains optima locaux peuvent se transformer et devenir *accessible* par d'autres états plus probables et ainsi perdre leur statut d'optima locaux.

Si l'espace de recherche S est fortement non-convexe relativement à la fonction de transition définie (S contient plusieurs optima locaux qui sont relativement profonds), il faut choisir une méthode de recherche capable de surmonter les optima locaux. On peut utiliser les méthodes stochastiques comme, par exemple, le recuit simulé (Geman et Geman, 1985) ou les algorithmes génétiques (Goldberg, 1989). Ici la transition vers des états moins probables est permise et contrôlée par un paramètre de température. La limitation principale de ces méthodes est leur temps de recherche.

Si S est convexe, on peut utiliser des méthodes de recherche déterministes qui convergent vers la solution globale. Dans ce cas, les méthodes descendantes comme, par exemple, le gradient descendant ou le gradient conjugué sont largement utilisées.

¹

Un espace de recherche est dit convexe lorsque pour toute état de cet espace il existe un nombre élevé de séries monotones de transitions distinctes commençant par cet état et *rejoignant* l'optimum global.

1.1.4 Organisation perceptuelle

La notion de l'organisation perceptuelle est inévitable dans un système de vision artificielle (Grimson, 1990; Horaud et Skordas, 1989; Lowe, 1985; McCafferty, 1993; Mohan et Nevatia, 1992; Sarkar et Boyer, 1993). Les scènes du monde réel sont constituées à partir d'organisations d'objets suivant des relations géométriques précises. De plus, chaque objet est généralement formé d'un assemblage de sous-objets. Finalement, chaque sous-objet peut être représenté par l'assemblage des surfaces qui délimitent son intérieur. Les images sont des projections 2-D des scènes et par conséquent, reflètent la réalité structurelle de cette scène.

L'organisation perceptuelle se traduit principalement par des regroupements de primitives extraites de l'image et du modèle. Les primitives de l'image sont généralement extraites par opérations de détection des discontinuités d'intensité. Ces opérations fournissent un ensemble fragmenté de primitives. Les regroupements de ces primitives permet de concentrer l'information et de produire des représentations denses et organisées afin de réduire l'espace de recherche de l'étape d'appariement. L'annexe B présente les règles généralement utilisées dans le regroupement des primitives. Dans un système de vision artificielle, la proximité, la similarité, la continuité et la symétrie se traduisent principalement par des contraintes géométriques souvent exprimées sous formes fonctionnelles ou algorithmiques. D'un autre côté, les primitives à regrouper doivent vérifier la propriété d'invariance qui est essentielle dans chaque procédure de reconnaissance et d'appariement (Marr, 1982). Elle se traduit par: tout appariement doit se faire entre deux entités, extraites respectivement de l'image et du modèle, issues de la même réalité physique. Par exemple, un appariement d'une paire de primitives extraite de deux images stéréo implique que les deux primitives correspondent à

la même réalité physique. L'invariance peut être définie par rapport aux perturbations dans les différents paramètres du processus de formation d'images: le point de vue, la position et le type de l'éclairage, la non-linéarité de la projection et le bruit.

Grimson (1990) utilise des méthodes de reconnaissance basées sur des techniques de recherche dans l'espace de correspondance entre le modèle et l'image. Il fait appel à des contraintes géométriques de type unitaire, binaire ou d'ordre supérieur. L'ordre indique le nombre de primitives impliquées dans chaque contrainte géométrique. Les contraintes expriment l'organisation perceptuelle entre le modèle et les primitives extraites de l'image. La recherche se fait par la construction d'un arbre ou d'un graphe des correspondances dans lequel chaque noeud représente une paire de primitives image-modèle. La similarité, la proximité et l'orientation sont souvent utilisées comme contraintes géométriques de type unitaire pour restreindre le nombre d'appariements. Les contraintes binaires ou d'ordre supérieur mesurent l'adéquation entre deux ou plusieurs primitives du modèle et les primitives correspondantes de l'image, et sont souvent basées sur la similarité, la continuité et la fermeture. L'utilisation de ces contraintes géométriques permet ainsi une réduction significative de la dimension de l'espace de recherche.

Les contraintes géométriques ont été utilisées également avant l'étape de l'appariement avec le modèle afin de regrouper les primitives dans des descriptions plus globales. McCafferty (1993) a utilisé une fonction d'objectif globale intégrant l'ensemble des contraintes géométriques. Ces contraintes traduisent les règles d'organisation perceptuelle présentées à l'annexe B. La recherche de la solution se fait par regroupement itératif des primitives afin d'optimiser la fonction d'objectif.

1.2 Problématique particulière à l'analyse des scènes semitransparentes

Après avoir examiné la problématique générale reliée à la vision artificielle, nous présentons dans cette section la problématique associée à l'analyse des scènes semitransparentes et plus particulièrement aux images médicales en radiographie de projection. Les résultats de notre système de détection et d'étiquetage doivent être intégrés dans une approche globale pour la reconstruction 3-D de la colonne vertébrale à partir d'un nombre limité d'images radiographiques. Rappelons que le présent travail se limite au développement du système de détection et d'étiquetage des structures anatomiques vertébrales apparaissant dans les radiographies du rachis. Afin de souligner l'utilité particulière de notre système de détection, nous discutons brièvement des méthodes de reconstruction 3-D.

Les méthodes de reconstruction 3-D en imagerie médicale peuvent être divisées d'une part en méthodes exactes et d'autre part en méthodes fonctionnant par inférence (de Guise *et al*, 1995). Pour les méthodes exactes (par exemple, rétroprojection filtrée et technique de reconstruction algébrique), on estime la valeur de la densité de la variable inspectée pour chaque pixel (ou voxel) du plan (ou du volume) à reconstruire. En tomographie axiale et en tomographie à émission de photon, la variable est respectivement l'atténuation photonique des tissus et la densité de radio-isotope capté par les tissus. Ces méthodes sont efficaces si l'ensemble des données observées est représentatif, c'est à dire lorsque le nombre des images observées est élevé et que l'angle de vue est réparti autour de la scène. Le problème devient sous-contraint lorsque le nombre d'images observées ou l'angle de vue sont très limités. Dans

ce cas, on doit régulariser le problème en faisant appel aux méthodes d'inférence de la vision artificielle qui introduisent des hypothèses sous forme de connaissances additionnelles (par exemple, modèles géométriques ou relationnels) de la scène à reconstruire (Hoffman *et al.*, 1988; Suetens *et al.*, 1988; Close, 1991). Cet apport d'information compense le nombre limité d'images observées. Contrairement aux méthodes exactes, les méthodes par inférence offrent généralement des résultats sous forme de représentation de frontières (contours ou surfaces). Ces techniques d'inférence sont toujours précédées d'un processus d'extraction et d'étiquetage qui est le centre d'intérêt de cette thèse.

La problématique de l'analyse des images médicales reflète les propriétés spécifiques des différentes modalités d'imagerie. Les problèmes rencontrés au niveau primaire dans l'analyse des images médicales sont généralement plus complexes que ceux rencontrés dans l'analyse des images de scènes naturelles². Cela est dû à des difficultés relevant du processus de formation d'image ou de la nature intrinsèque des scènes médicales.

1.2.1 Difficultés liées au processus de formation d'image

La discussion portera sur les difficultés liées au processus de formation d'image: le rapport signal sur bruit (S/B), la résolution, la semitransparence, la superposition des structures et le contraste.

Les images médicales sont caractérisées par un faible S/B et une faible *résolution spatiale* si on les compare, par exemple, aux images photographiques. Pour les modalités d'imagerie

² Ce sont des scènes de la vie courante où les objets sont habituellement opaques et le processus d'acquisition d'images est photographique.

de transmission ou d'émission, cela est directement relié à la nature statistique des processus photoniques (poissonniens) et aux caractéristiques des capteurs. En radiographie conventionnelle, la densité surfacique moyenne de photons reçus par le capteur ou le film est de l'ordre de 10^7 photons/cm², elle est réduite à 10^2 photons/cm² en médecine nucléaire, alors que la densité en photographie conventionnelle est de l'ordre de 10^{12} photons/cm². Un faible rapport S/B se traduit pratiquement par une sur-fragmentation des primitives extraites. Cela a pour effet d'augmenter la dimension de l'espace de recherche et l'incertitude relative au processus de recherche et de réduire la stabilité de la technique de recherche. D'autres facteurs liés au processus de formation d'image et spécifiques à chaque modalité d'imagerie viennent aussi nuire à la qualité d'image, comme l'atténuation, la diffusion, la diffraction, la défocalisation, l'échantillonnage, l'interférence inter-tranche, les paramètres ambiants, et le bruit de l'instrumentation.

L'opacité des surfaces par rapport à la lumière visible est une hypothèse implicite utilisée dans les systèmes d'analyse d'images photographiques (Fisher, 1989; Grimson, 1981; Horn, 1986; Gibson, 1950). En imagerie médicale, différents types de signaux³ sont utilisés afin de rendre l'intérieur du corps humain visible. La nature du signal utilisé et le processus d'acquisition d'image ne sont pas les mêmes en imagerie médicale et en photographie. Ainsi, on peut définir les *scènes semitransparentes* (SST) par le processus de formation d'images dont le signal d'exploration a une capacité pénétrante de la matière qui à son tour atténue ce signal. Habituellement, la distribution spatiale de l'atténuation constitue le contenu

3

Par exemple l'onde électromagnétique (photonique) pour l'imagerie à transmission (rayon-X), l'imagerie à émission (rayon γ) et l'imagerie à résonance magnétique (radio fréquence).

informationnel des images médicales. Un système de vision artificielle pour l'analyse des SST doit inévitablement prendre en considération la propriété de semitransparence pour mieux représenter la réalité de la scène.

L'opacité d'objets crée le problème de l'occlusion dans l'analyse de scènes naturelles. Ce problème n'est pas défini de la même façon pour l'analyse de SST où les objets ne s'occluent pas mais se superposent. Cela est désigné par le problème de la *superposition des structures*. Les projections des scènes semitransparentes créent ainsi des images très complexes sur des fonds eux aussi complexes et variables. Il y a aussi le problème de *variation du contraste*. Dans des projections radiographiques, l'intensité de chaque pixel de l'image résulte de l'intégration linéaire de l'atténuation photonique suivant la direction de la projection. Par conséquent, le contraste dépend de l'épaisseur des structures qui se superposent et des distributions intrinsèques de l'atténuation photonique.

1.2.2 Difficultés reliées à la nature des scènes

Dans cette section, la discussion portera sur trois difficultés reliées à la nature des scènes médicales: la *non-rigidité* des objets, la géométrie *non-polyédrique* et la *complexité* des formes.

La rigidité des objets est une propriété souvent utilisée comme hypothèse inhérente en vision artificielle afin d'introduire des contraintes géométriques dérivant des principes de l'organisation perceptuelle. En effet, cette hypothèse est essentielle pour plusieurs méthodes d'inférence de la forme 3-D à partir de quelques images 2-D (Grimson, 1990; Lowe, 1987; Ullman, 1979a; Huttenlocher et Ullman, 1987; Jacobs, 1992; Brooks, 1981, 1986).

Cependant, les objets formant les scènes en imagerie médicale ne sont *pas rigides* généralement, ce qui prive le système d'analyse d'image d'une contrainte importante. Malgré le fait que nous soyons intéressés par la détection des structures osseuses qui sont considérées mécaniquement solide, la variation⁴ dans la forme géométrique d'une structure osseuse peut être désignée comme la *non rigidité* de la forme. Toutefois, chaque objet des scènes médicales possède sa propre structure géométrique qui identifie et décrit sa forme. En fait, nous pensons que l'information de la structure spécifique à chaque objet est utilisée par le processus de reconnaissance de la vision humaine (Witkin et Tenenbaum, 1983). Quoique les représentations géométriques et relationnelles des objets soient discutées et utilisées dans cette thèse, le sujet de la représentation optimale ne constitue pas le centre de cette recherche (Jacobs, 1992).

L'importance des discontinuités des surfaces dans un système de vision artificielle provient de leur propriété d'*invariance* par rapport aux paramètres du processus de formation d'image. Les discontinuités de surface de premier ordre sont les invariants les plus utilisés en vision artificielle (Marr, 1982). En général, les discontinuités des surfaces se projettent dans les images sous forme de discontinuités d'intensité. En effet, la plupart des techniques d'analyse de niveau primaire identifient les invariants géométriques des objets en détectant les discontinuités d'intensité dans l'image. Les organes du corps humain ont généralement des formes géométriques *non-polyédriques* à surfaces arrondies avec de faibles⁵ discontinuités de premier ordre. Cela crée des images à *faible contraste*.

⁴ Due à une multitude de paramètres pathologiques ou naturels.

⁵ Relativement aux objets polyédriques rencontrés dans les scènes naturelles.

Enfin, les objets de scènes médicales ont habituellement des formes géométriques *complexes*. Cela est également vrai pour les différentes façons pour décrire un objet comme, par exemple, la structure, le squelette, les surfaces et les propriétés locales des tissus (couleur, diffraction, réflexion, atténuation, ...). La complexité de ces objets se traduit donc par des images complexes à analyser.

1.3 Objectifs et hypothèses

L'objectif général de ce travail de recherche est de développer un système de vision artificielle pour la reconnaissance automatique d'objets dans des images de scènes semitransparentes. Dans les sections 1.1 et 1.2, nous avons discuté respectivement de la problématique générale de système de vision artificielle et des difficultés liées aux scènes semitransparentes. Les objectifs spécifiques de ce système sont les suivants:

1. délimiter l'objet à reconnaître et ses principales composantes (sous-objets);
2. détecter les frontières de l'objet avec précision et robustesse en tenant compte de la nature déformable de l'objet;
3. étiqueter les différentes régions de l'image contenant l'objet et ses composantes.

Sur le plan méthodologique, puisque l'implantation d'un système de vision peut être présentée comme un problème de recherche dans un espace d'état, deux autres objectifs sont énoncés concernant la formulation du problème :

4. réduire la dimension de l'espace de recherche;
5. trouver une solution optimale.

La reconnaissance d'objet consiste à identifier l'objet et à déterminer sa pose en appariant les images observées avec différents modèles. Dans la présente thèse, on ne s'intéresse pas à la recherche dans l'espace des poses. Le travail se concentre uniquement sur le problème d'*appariement* d'un *modèle déformable* connu *a priori* avec l'image observée. Nous supposons que l'identité de l'objet est connue et que la *pose* peut être *approximativement estimée* par un processus externe de calcul. Cela nous permet de réduire la dimension de l'espace de recherche ainsi que la complexité du problème. Cependant, l'estimation de la pose sera utilisée pour proposer une projection 2-D à partir d'un modèle 3-D de l'objet à détecter. Nous supposons également que si l'incertitude dans l'estimation de la pose est élevée, l'étape de l'estimation pourrait alors proposer plusieurs projections 2-D du modèle. En effet, cet ensemble de projections représenterait différents échantillons du graphe d'aspect du modèle (Dickinson *et al.*, 1990, 1992). De plus, cela pourrait aussi représenter la tolérance et le niveau de déformation du modèle.

On suppose aussi que l'hypothèse de la *non-accidentalité*⁶ est vérifiée. Cela signifie que le point de vue dont est issue l'image ne peut pas causer une dégénérescence⁷ de la forme géométrique du modèle 2-D qui représente l'objet en projection. De plus, cette hypothèse stipule que les modèles utilisés représentent fidèlement les objets de la scène. Donc, les images vérifiant l'hypothèse de la non-accidentalité sont issues d'un point de vue général et d'un arrangement général des objets de la scène tel que les modèles utilisés correspondent à

⁶ Généralement utilisée dans plusieurs travaux de la vision artificielle.

⁷ Par exemple, si le point de la projection appartient à un des plans définis par les surfaces planes qui délimitent les objets de la scène, la projection de cette surface dans le plan de l'image se dégénère en une droite.

la réalité existante dans l'image (Witkin et Tenenbaum, 1983). La *non-accidentalité* a été également proposée pour résumer les règles de l'organisation perceptuelle (Annexe B). D'ailleurs, cela s'exprime par l'énoncé suivant: un arrangement spécifique⁸ des primitives extraites de l'image ne peut pas être dû à une pose ou à une position accidentelle mais reflète l'organisation réelle de la scène.

On rappelle que les objets d'images médicales ont une géométrie complexe. On suppose que la géométrie de chaque objet peut être modélisée par l'assemblage de plusieurs sous-objets suivant un graphe de connexion dit aussi *graphe relationnel intrinsèque*. Cela est désigné par l'hypothèse de la *géométrie structurelle*.

On rappelle aussi que l'application précise de l'approche développée dans cette thèse est la reconnaissance des structures osseuses⁹ à partir d'images radiographiques. Conscient de la grande difficulté qu'un tel système de reconnaissance doit surmonter à cause de l'imbrication des objets, de leurs structures déformables et complexes et de la semitransparence de la projection, nous avons choisi d'introduire une hypothèse simplificatrice dite de l'*objet isolé*. En effet, nous proposons d'aborder le problème en supposant que chaque image radiographique à analyser contient un seul objet isolé. Malgré le fait que cette hypothèse ne corresponde pas à la réalité de la colonne vertébrale, nous pensons cependant qu'aborder le problème dans ce sens constitue une étape pour proposer ultérieurement un système complet d'analyse d'images radiographiques de la colonne vertébrale.

⁸ Suivant les règles de l'organisation perceptuelle.

⁹ Plus particulièrement, les vertèbres de la colonne vertébrale.

Les hypothèses du *graphe relationnel intrinsèque* et de *l'estimation de la pose* justifient l'introduction de l'hypothèse de la *préservation du modèle*. Cela se traduit par une règle de préservation de la structure relationnelle ou *topologique* de la solution durant le processus de recherche: cette structure doit rester constamment identique à celle du modèle initial.

1.4 La thèse

Ce premier chapitre a été consacré principalement à la présentation du cadre général de la recherche. Nous avons d'abord énoncé les objectifs ultimes de notre recherche et les objectifs particuliers du présent travail. Ensuite, nous avons présenté et discuté la problématique générale d'un système de vision artificielle et la problématique particulière reliée au système d'analyse des SST. Précisément, nous avons rappelé que le problème inverse en vision artificielle est mal posé et sous-contraint. Nous avons aussi présenté et discuté le problème de la recherche dans l'espace des états possibles. Puis, nous avons souligné l'importance des contraintes géométriques découlant des principes de l'organisation perceptuelle dans la stabilisation du problème. La problématique particulière de l'analyse des images médicales a été discutée d'un point de vue théorique avec la présentation des caractéristiques des scènes semitransparentes retrouvées en radiographie de projection. Finalement, nous avons présenté les hypothèses générales et simplificatrices de notre approche.

La suite de la thèse est organisée en cinq autres chapitres. Le chapitre 2 résume l'approche d'une façon intégrale. Il introduit la nouvelle approche de *reconnaissance par appariement avec un modèle*. Le modèle 3-D et la formation du modèle 2-D dans le plan de

l'image sont brièvement décrits ainsi que la nouvelle technique d'alignement global et la nouvelle technique d'appariement par déformation progressive du modèle.

La nouvelle technique d'alignement global du modèle est présentée en détail au chapitre 3 (article 1). Cette technique est guidée par un critère probabiliste. La fonction du critère a une forme distribuée et établit un champ d'interaction entre le modèle et l'image. Ce champ est interprété comme un champ de force appliqué au modèle. Sous l'effet de cette force, le modèle subit un ensemble de transformations. La solution est atteinte par intégration temporelle des équations dynamiques du système.

L'étape qui suit l'alignement global du modèle est l'appariement du modèle avec l'image par déformation progressive.

Le chapitre 4 (article 2) permet d'abord d'introduire le modèle *simple* à multiples courbes. Ce chapitre est consacré à la formulation stochastique. Nous utilisons un nouveau critère stochastique appelé la *pseudo-probabilité conjointe*. Ce critère décrit l'interaction entre l'image observée et les courbes du modèle 2-D. L'approche est validée sur des images synthétiques fortement bruitées et des modèles planaires à multiples courbes.

Au chapitre 5, nous développons une nouvelle version du modèle à multiples simples courbes fermées pour représenter en 2-D des projections semitransparentes et non-accidentelles d'objets volumétriques. On modifie le critère probabiliste afin de tenir compte de cette nouvelle représentation. Ce chapitre se concentre sur la modélisation par simples courbes fermées et sur la dualité ligne-région. La dernière partie du chapitre 5 détaille la technique de recherche (d'appariement) par déformations locales et progressives. La méthode de recherche est descendante et à échantillonnage aléatoire. On explique le processus

d'échantillonnage et le critère d'acceptation. Chaque nouvel échantillon est soumis à une contrainte fonctionnelle et à des contraintes algorithmiques. Ces dernières découlent directement de la préservation de l'aspect topologique du modèle. Cette version est par la suite testée avec des images synthétiques et des images réelles d'un objet isolé.

La synthèse de l'approche et les voies futures de recherche sont présentées au chapitre 6.

Chapitre 2

Présentation générale de l'approche

2.1 Introduction

Ce chapitre présente une vision globale de l'approche qui sera détaillée dans les trois prochains chapitres et qui est intitulée: *Modèle probabiliste de courbes déformables pour appariement avec images de scène semitransparente*.

On rappelle que l'objectif ultime de notre système de vision est de reconnaître avec précision les différentes structures de la colonne vertébrale sur des images radiographiques. La *figure 2.1* montre une

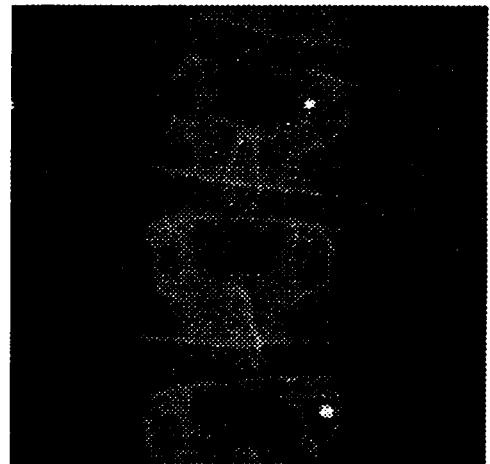


Figure 2.1. Exemple de projection radiographique postéro-antérieure de vertèbres humaines.

portion de projection radiographique postéro-antérieure (PA0°) de la colonne vertébrale d'un sujet scoliotique. L'analyse de ce type d'image comporte des difficultés particulières comparativement à l'analyse des scènes opaques. Ces difficultés peuvent être reliées au processus de formation d'image (la superposition des structures, le faible S/B, et le contraste faible et variable), à la géométrie des objets et à l'organisation de la scène (les objets sont

non-polyédriques, complexes et déformables et leur imbrication augmente davantage la complexité des images). La *figure 2.2* montre les bords extraits de l'image de la *figure 2.1* à l'aide d'un opérateur de détection de discontinuités. Notons la fragmentation des éléments de frontières causée par les difficultés mentionnées plus haut.

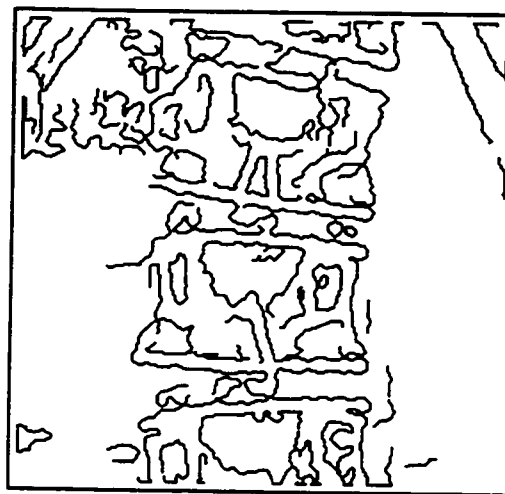


Figure 2.2. Résultat de la détection des discontinuités de la *figure 2.1*. Le passage par zéro de la convolution avec le Laplacien d'une Gaussienne $N(\mu=0, \sigma=2)$ a été masqué par le gradient binarisé à un seuil de 0.2.

Notre approche de détection et d'étiquetage est basée sur l'appariement d'un modèle 2-D déformable et connu *a priori* avec l'image observée.

On suppose que les hypothèses avancées au premier chapitre s'appliquent au modèle et à l'image à analyser. L'hypothèse de l'*objet isolé* est introduite uniquement dans le but de simplifier partiellement les images à traiter. La *figure 2.3* montre une image bruitée de projection radiographique¹⁰ d'un objet synthétique immergé dans une cuve remplie d'eau et schématisant la géométrie et les propriétés d'atténuation d'une vertèbre isolée. Le résultat de la détection des discontinuités est également illustré. Les caractéristiques de cet objet ainsi que les conditions de la création de cette image sont données au chapitre 5.

¹⁰

Issue d'une pose générale non-accidentelle.

L'approche est formée de deux parties: la modélisation et la recherche de la solution. La modélisation consiste à trouver une représentation géométrique correspondant aux objets à reconnaître. Nous avons choisi une

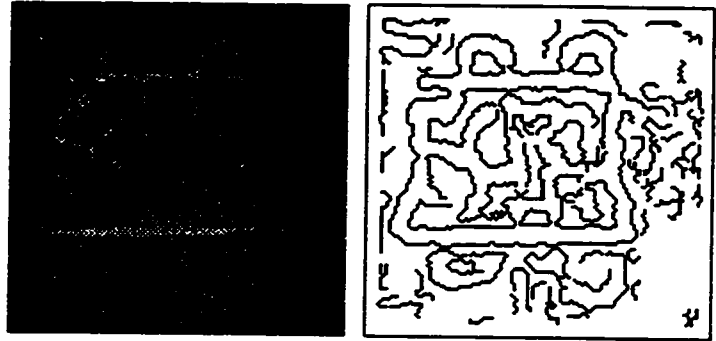


Figure 2.3. Projection radiographique d'un objet isolé et détection des discontinuités utilisant les mêmes paramètres qu'à *figure 2.2*.

modélisation structurelle des objets 3-D en assemblant trois sortes de sous-objets simples. Le modèle 2-D utilisé pour l'appariement avec l'image résulte de la projection du modèle 3-D dans le plan de l'image à partir d'un point de vue non-accidentel supposé connu approximativement. Le modèle 2-D peut être décrit par la superposition d'un ensemble de simples courbes fermées et désigné par modèle à multiples courbes (MMC). Cette partie est détaillée dans la section 2.2.

Pour éviter les méthodes stochastiques de recherche (Geman et Geman, 1985), coûteuses en temps de calcul, on propose une méthode descendante à deux étapes: l'alignement global¹¹ et l'appariement par déformation progressive¹². Pour améliorer la performance de la technique d'appariement par déformation, on l'intègre dans une structure pyramidale de multirésolution tout en la précédant d'une nouvelle technique d'ajustement (ou d'alignement) global de la solution initiale (section 2.3).

L'appariement consiste à déformer le modèle géométrique 2-D¹³ afin d'atteindre une

¹¹ Développé au chapitre 3.

¹² Développé aux chapitres 4 et 5.

¹³ Appelé aussi solution initiale.

solution optimale. L'appariement du MMC avec l'image est guidé par l'optimisation d'une fonction d'objectif développée à partir d'un critère probabiliste et par des contraintes algorithmiques afin de préserver le contenu topologique du modèle. Cette méthode est détaillée dans la section 2.4.

2.2 Modèle géométrique

Le processus de reconnaissance d'objets est souvent basé sur un modèle. Il est habituellement itératif et formé d'une étape d'identification de l'objet et d'une étape d'association et de cumul de preuves¹⁴. Cette section présente une revue de différents modèles géométriques utilisés en analyse d'images suivie d'une description du modèle que nous proposons pour l'appariement avec images de SST. Quatre aspects ont influencé le choix de modèle: le *contenu* informationnel des images, la *non-rigidité* des objets, l'*organisation perceptuelle* et la *simplicité* de la représentation.

2.2.1 Revue des modèles géométriques

Les modèles rigides étaient souvent utilisés en vision artificielle¹⁵ (Ullman, 1979; Yuille et Ullman, 1989). Ils impliquaient l'imposition de contraintes physiques dérivant de la nature solide des objets. Toutefois, notre travail vise l'analyse des images d'objets déformables. Il faut donc envisager des modèles capables de tenir compte de cette nature déformable.

¹⁴ Apparier les attributs du modèle avec les primitives extraites de l'image afin de confirmer ou de refuser l'identité proposée.

¹⁵ Dans l'inférence de la structure ou du mouvement à partir d'une séquence d'image, la reconnaissance d'objet, l'inférence de la pose à partir d'une ou plusieurs images.

Les méthodes basées sur la description et l'appariement de silhouette, comme, par exemple, les descripteurs de Fourier (Wallace, 1980; Reeves *et al.*, 1988) et la description dans l'espace des échelles (Mokhtarian et Mackworth, 1992), sont sensibles au choix des méthodes d'extraction des contours¹⁶. L'efficacité de ces représentations est largement compromise par la sur-fragmentation des éléments de discontinuité. D'autres méthodes ont été développées pour la reconnaissance d'objets déformables et allongés (Subirana-Vilanova, 1993) ou d'objets en forme d'arbre (réseau coronarien dans des vues radioscopiques) (Close, 1991; Hoffman *et al.*, 1988) et sont basées sur l'appariement d'un modèle de squelette¹⁷ avec le squelette extrait de l'image. L'information structurelle du squelette s'est avérée importante dans le processus de reconnaissance de la vision humaine (Marr, 1982; Witkin et Tenenbaum, 1983). Cependant, afin de pouvoir extraire le squelette, ces méthodes nécessitent la connaissance de la silhouette de l'objet à détecter.

Les modèles élastiques sont utilisés dans plusieurs applications comme l'appariement des connaissances ou la mise en correspondance des cartes anatomiques¹⁸ (Bajscy et Kovacic, 1989; Feldmar et Ayache, 1996). Au chapitre 5, nous discutons de ces modèles élastiques et les comparons à celui de notre approche. L'annexe C présente les aspects théoriques reliés à ces modèles. Les modèles élastiques sont principalement limités par leur capacité de déformation qui dépend d'un compromis entre la robustesse et la précision. Lorsque les

¹⁶ Ces méthodes supposent qu'elles disposent de courbes continues.

¹⁷ Le squelette peut être défini en vision artificielle comme le croquis de lignes qui résume la représentation canonique de la forme (Marr, 1982; Subirana-Vilanova, 1993). Plusieurs définitions mathématiques d'un squelette ont été proposées dans la littérature. Subirana-Vilanova (1993) présente une revue des techniques existantes pour l'extraction de squelette.

¹⁸ "Knowledge registration or atlas matching."

images sont fortement bruitées et les objets hautement déformables, ces modèles ne garantissent pas la précision de la détection.

Au cours des quinze dernières années, les contours dynamiques¹⁹ (CD) ont été largement utilisés en vision artificielle (Curwen et Blake, 1992; Dubuisson et Jain, 1995; Geiger *et al.*, 1995; Leymarie et Levine, 1993) grâce à leurs propriétés inhérentes²⁰ permettant de surmonter deux difficultés de taille dans l'organisation perceptuelle: le regroupement des fragments de discontinuités d'intensité et la fusion des régions. L'ajustement des CD à l'image est souvent guidé par l'optimisation d'une fonction d'objectif donnée par la somme de deux termes²¹ (Terzopoulos, 1987a;b). Le premier terme représente la contrainte de lissage souvent donnée sous forme de stabilisateur quadratique (voir l'annexe A). Le deuxième terme représente la force externe exercée par l'image sur les contours. Ce terme peut aussi incorporer une force d'inflation ou de déflation²² (Cohen, 1991; Kapur, 1995). D'autres modèles paramétriques peuvent intégrer (Jain *et al.*, 1996; Lai et Chin, 1995; Staib et Duncan, 1992) ou non (Bardinet *et al.*, 1996) un troisième terme de biais vers une distribution *a priori* des paramètres. Cela compromet la précision de détection et limite la flexibilité de la déformation du modèle. Les CD peuvent aussi être combinés à différentes techniques de segmentation d'images (Zhu et Yuille, 1996; Kapur, 1995). Jusqu'à présent, l'utilisation de ces CD s'est limitée à des modèles à contour unique ou à plusieurs contours disjoints.

¹⁹ Un contour dynamique est une courbe plane, continue, fermée et ne se croisant pas et dont l'intérieur ne contient pas de la matière par comparaison avec les modèles élastiques.

²⁰ Connectivité et fermeture.

²¹ Cette forme est désignée par "snake."

²² Cette forme est désignée par "balloon."

2.2.2 Modèle de multiples simples courbes fermées

Nous proposons un modèle représenté théoriquement par un assemblage structurel d'objets élémentaires 3-D désignés par le terme de *sous-objets* (Marr, 1982; Pentland et Horowitz, 1991). Cette représentation riche et *simple* convient à la modélisation des vertèbres humaines. Les sous-objets sont décrits uniquement par leurs surfaces délimitantes. Cette représentation de frontières convient à la nature semitransparente des scènes et aussi au *contenu informationnel* des images qui réside essentiellement dans les frontières et par conséquent dans les discontinuités d'intensité. Nous nous limitons à un ensemble de sous-objets à trois formes libres²³ planaires, coniques et cylindriques. À partir d'une pose connue²⁴, la projection semitransparente d'un sous-objet peut ainsi être décrite par une simple courbe fermée extrémale pouvant intégrer une ou deux autres simples courbes fermées internes²⁵. Le terme "simple courbe fermée" (SCF) est équivalent au terme "contour dynamique". L'avantage de l'utilisation de SCF dans le modèle géométrique découle principalement de sa représentation *déformable* et de ses propriétés intrinsèques directement reliées à l'aspect de *l'organisation perceptuelle*. La connexité et la fermeture de SCF réunissent implicitement les propriétés de proximité, continuité, fermeture et dualité ligne-région qui sont habituellement utilisées sous forme de critères de regroupement dans *l'organisation perceptuelle* des primitives extraites de l'image (annexe B). La dualité ligne-région est importante car elle permet d'établir une liaison entre le niveau structurel du modèle et le niveau primaire dans un système de vision.

²³ La désignation des *formes libres* est en opposition avec les formes paramétriques et les formes élastiques.

²⁴ L'hypothèse de la *pose approximativement estimée* par un processus externe.

²⁵ Ici, on fait appel à l'hypothèse de la *non-accidentalité* afin de limiter la représentation à cet ensemble de simples courbes fermées.

Le modèle de frontières utilisé dans l'appariement résulte de la projection du modèle 3-D dans le plan de l'image. Le modèle 2-D est dénommé *modèle à multiples simples courbes fermées* pour *formes libres*. Le modèle utilisé dans le chapitre 4 est *simple* car il ne contient que des SCF extrémales. Le modèle *complet* est plus réaliste et intègre en plus des SCF internes (chapitre 5).

2.3 Alignement global du modèle

L'étape d'alignement global est importante pour les méthodes descendantes de recherche car elle contribue à augmenter les possibilités de convergence vers une solution optimale. L'effet de cet alignement se traduit par un rapprochement de la solution initiale vers la solution optimale. On désire que la technique puisse s'appliquer directement à l'image d'intensité et qu'elle soit robuste et invariante aux paramètres du processus de formation d'image.

2.3.1 Revue des méthodes d'alignement global

Plusieurs méthodes d'alignement global ont été proposées pour différentes applications en vision artificielle: le positionnement, la reconnaissance, la poursuite et l'appariement. Ces méthodes sont directes ou itératives. Les méthodes directes sont basées sur les propriétés d'invariance de certaines transformations comme, par exemple, les moments d'inertie (Reeves *et al.*, 1988; Hu, 1962) ou les descripteurs de Fourier (Wallace et Wintz, 1980). Les méthodes itératives sont souvent basées sur l'optimisation d'une fonction d'objectif qui mesure l'adéquation entre le modèle transformé et l'image observée (Lavallée et Szeiliski,

1995; Feldmar et Ayache, 1996).

Les méthodes d'alignement global peuvent être basées soit sur l'intensité (de l'image) soit sur les primitives (extraites de l'image). Les méthodes basées sur l'intensité sont généralement très sensibles aux variations d'intensité du fond de l'image ainsi qu'aux variations des paramètres de formation d'image. De plus, ces méthodes ne sont pas efficaces avec des alignements de grande disparité et avec des images bruitées ou sous-échantillonnées. Par ailleurs, les méthodes basées sur les primitives sont généralement précédées d'une étape d'extraction et d'une étape de regroupement, ce qui les rend sensibles aux résultats de ces deux étapes. Cette sensibilité est élevée pour les méthodes utilisant l'appariement point-à-point (Feldmar et Ayache, 1996). Le chapitre 3 présente une revue bibliographique détaillée des méthodes d'alignement global.

2.3.2 Modèle dynamique pour alignement global

On présente, au chapitre 3, une nouvelle technique d'alignement global. Cette technique permet l'alignement itératif d'un modèle 2-D avec l'image observée en utilisant les transformations suivantes: translation, rotation, mise à l'échelle dans deux directions non-orthogonales et cisaillement. Les grandes lignes de cette technique sont les suivantes:

1. l'alignement se fait par optimisation itérative d'une fonction d'objectif;
2. la fonction d'objectif est développée à partir d'un critère probabiliste à forme distribuée.

Elle exprime une mesure locale de proximité et de similarité entre le modèle et le gradient de l'image;

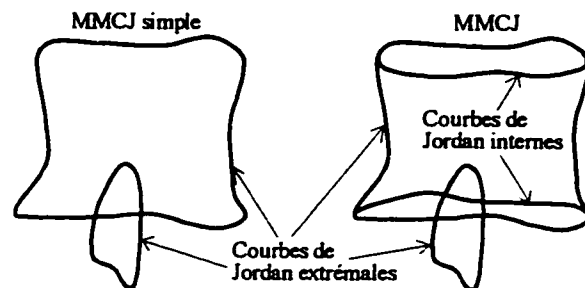
3. le processus de recherche est modélisé par un système dynamique. La solution est donnée

par intégration temporelle des équations différentielles de ce système. Les transformations sont organisées dans une structure hiérarchique qui améliore la performance de la technique.

La technique peut s'appliquer à différents types d'images avec des modèles à lignes (multiples courbes) et des modèles à intensité (images). La technique est stable et robuste et elle converge bien même en présence de grandes disparités entre modèle et image.

2.4 Appariement du modèle par déformation

L'étape qui suit l'alignement global est la déformation du modèle afin d'atteindre un appariement optimal entre le modèle aligné et l'image. La



technique de déformation du modèle est aussi appelée la technique de recherche. Elle est guidée par un critère probabiliste menant à l'optimisation d'une fonction d'objectif. En premier lieu, on développe le critère pour un MMC *simple* contenant uniquement des SCF extrémales (figure 2.4). Le chapitre 4 est consacré à la représentation du MMC *simple*, au choix du critère probabiliste, ainsi qu'à l'énoncé et à la justification des hypothèses et des approximations utilisées pour développer ce critère. Le modèle est réexaminé au chapitre 5 dans un cadre topologique pour une nouvelle représentation (plus réaliste) de MMC avec des SCF extrémales et internes.

Figure 2.4. Modèle *simple* de multiples SCF (MMC *simple*) à deux courbes extrémales (à gauche). MMC *complet* pour représentation de projections d'objets volumétriques à deux courbes extrémales et deux courbes internes (à droite).

2.4.1 Revue des méthodes de recherche

Les méthodes de recherche utilisées pour apparier modèle et image peuvent être de deux types: (i) les méthodes oeuvrant dans l'espace de correspondance des traits caractéristiques convenant à la reconnaissance d'objets rigides (Grimson, 1990; Lowe, 1985, 1987; Horaud et Skordas, 1989; Brooks, 1981, 1986); et (ii) les méthodes d'optimisation de critère global pouvant être utilisées tant pour les objets rigides que pour les objets non-rigides (Terzopoulos, 1987a, b; Bajscy et Kovacic, 1989; Storvik, 1994).

Les méthodes d'optimisation peuvent être déterministes ou stochastiques. Les méthodes stochastiques (recuit simulé, équations différentielles stochastiques, algorithmes génétiques) peuvent assurer la convergence du processus d'optimisation vers une solution globale mais à coût très élevé en terme de temps de recherche (Geman et Geman, 1985; Grenander et Miller, 1994; Amit *et al.*, 1991; Goldberg, 1989; Storvik, 1994). Les méthodes déterministes (descentes, à gradient, variationnelles, programmation dynamique) sont souvent utilisées pour des problèmes convexes (Terzopoulos, 1987a, b; Bajscy et Kovacic, 1989; Jain *et al.*, 1996; Lai et Chin, 1995; Staib et Duncan, 1992). Cependant, ces méthodes pourraient servir à résoudre des problèmes non-convexes lorsque dotées de fonctions de transition efficaces (voir les discussions à la section 1.1.3) et précédées d'étapes d'alignement global.

2.4.2 Appariement par déformation aléatoire

La technique d'appariement développée dans cette thèse est descendante. Elle consiste à déformer progressivement et d'une façon aléatoire le MMC. La recherche est assujettie à la

contrainte fonctionnelle dérivant d'un critère probabiliste, et des contraintes algorithmiques imposées par l'hypothèse de *préservation topologique* du modèle. Une déformation est ainsi retenue si elle vérifie toutes les contraintes. L'étape de déformation est incorporée dans une structure de multirésolution (Terzopoulos, 1986a) qui améliore la capacité de surmonter les optima locaux en introduisant des interactions à grande échelle. Au début, la recherche par déformation est effectuée au niveau de résolution le plus grossier. La solution est ensuite projetée dans le niveau adjacent puis raffinée par la même technique de déformation. La dernière étape est répétée dans la même direction jusqu'à la résolution la plus élevée. La technique de recherche par déformation aléatoire est détaillée au chapitre 5.

2.4.3 Critère probabiliste

Notre approche d'appariement est développée dans un cadre stochastique, ce qui permet de choisir un critère probabiliste à propriétés intrinsèques, d'introduire des hypothèses précises dans le développement du critère et de les ajuster en fonction du résultat recherché. Il est souhaitable que la fonction d'objectif du critère ait une forme distribuée et puisse intégrer une mesure d'adéquation entre la solution, d'un côté, et l'image observée et le modèle initial, de l'autre. La solution produite doit être précise et robuste par rapport au bruit. Au chapitre 4, on discute des critères probabilistes les plus utilisés en analyse d'image. La maximisation de la *pseudo-probabilité conjointe* (MPPC) est introduite dans ce travail comme critère d'appariement offrant les propriétés souhaitables ci-haut mentionnées.

Soit I le modèle à appairer avec l'image observée Y . I et Y sont tous les deux définis sur la même grille de quantification L . La solution du MPPC est donné par:

$$l^* = \arg \max_l \left\{ \prod_{s \in L} p(l_{\delta s}, Y_{\delta s}) \right\} \quad (2.1)$$

où s désigne un site appartenant à L et δs est un système de voisinage centré sur s . En utilisant la dualité ligne-région des SCF avec des approximations convenables, la fonction d'objectif est donnée par la somme de trois termes. (i) Le terme relié aux données qui mesure l'adéquation de la solution l par rapport à l'image Y et assure la coexistence des SCF dans le contexte de la détection précise et robuste. (ii) Le terme relié au modèle qui mesure l'adéquation entre la solution et les connaissances topologiques du modèle et introduit ainsi une interaction entre le niveau structurel du modèle et le niveau de la détection. (iii) Le terme de lissage qui est donné par la somme des longueurs et des variations quadratiques des premier et deuxième ordres de la direction locale des SCF.

Le critère de la MPPC est développé au chapitre 4 pour la représentation *simple*²⁶ du MMC. Le critère est réajusté au chapitre 5 pour une représentation *complète*²⁷ du MMC. La technique de recherche est détaillée au chapitre 5.

L'optimisation de la fonction d'objectif est appelée la contrainte fonctionnelle imposée au processus de recherche. La méthode de recherche utilisée est descendante et déforme itérativement le modèle en respectant la contrainte fonctionnelle.

2.4.4 Contraintes algorithmiques

En plus de la contrainte fonctionnelle imposée par l'optimisation du critère global, le

²⁶ Contient seulement des SCF extrémales.

²⁷ Contient des SCF extrémales et des SCF internes.

processus de recherche (ou de déformation) est soumis à des contraintes algorithmiques. Ces contraintes sont issues de deux principes: (i) La préservation des SCF, et (ii) la préservation du contenu relationnel de l'état initial du modèle.

Le premier principe se traduit par les contraintes suivantes: la déformation doit préserver la continuité, la fermeture et l'orientation de la SCF et doit aussi empêcher les croisements d'une même SCF. Ces contraintes sont intégrées dans l'étape de la génération de nouveau segment de courbe.

On montre, au chapitre 5, que le deuxième principe se traduit par deux contraintes (ou conditions): s'assurer que la description syntaxique séquentielle du nouveau segment de courbe est identique à celle du segment à remplacer; et vérifier que l'ensemble des étiquettes de régions touchées par la déformation est inclus dans l'ensemble des étiquettes de la description syntaxique du segment déformé.

2.5 Conclusion

L'approche faisant l'objet de cette thèse est basée sur un modèle. On utilise les propriétés générales de la scène et du processus de formation d'image pour proposer un modèle déformable à multiples SCF. Le modèle est formé d'un ensemble de SCF extrémales et internes. Chaque courbe extrémale peut contenir dans sa région interne jusqu'à deux courbes internes disjointes. La superposition des SCF représente la semitransparence de la projection.

La détection et l'étiquetage se font par appariement du modèle avec l'image en deux étapes: l'alignement global et l'appariement par déformation progressive. Chacune des étapes

est itérative et guidée par l'optimisation d'une fonction d'objectif développée dans un cadre probabiliste. L'étape de l'alignement global utilise une technique dynamique de recherche. L'appariement se fait par une technique descendante de déformation progressive et aléatoire de la solution. La déformation est aussi soumise à des contraintes algorithmiques représentant le principe de préservation de modèle.

Chapitre 3

Alignement global du modèle

Article: *Dynamic model for global alignment and scaling*

Notre approche de détection et d'étiquetage est basée sur l'appariement d'un modèle déformable avec l'image observée. Une étape d'alignement global du modèle précède généralement l'étape de déformation. Cet article présente une nouvelle technique d'alignement global (Mallouche *et al.*, 1997a).

3.0 Résumé

L'étape d'alignement global est importante pour les méthodes descendantes de recherche. En effet, ces méthodes risquent de converger vers des optima locaux dans un espace de recherche non convexe si la solution initiale est relativement éloignée de la solution globale. Les grandes lignes de la nouvelle technique d'alignement global sont les suivantes:

1. l'approche est intégrée dans un cadre stochastique utilisant un critère global pour guider l'alignement;
2. l'interaction entre le modèle M et l'image observée Y comprend une mesure de similarité de l'orientation et une mesure de proximité;
3. les transformations globales sont une translation, une rotation, une mise à l'échelle et un

cisaillement;

4. la méthode de recherche est du second ordre et organisée dans une procédure hiérarchique.

Dans une première étape, on choisit la *maximisation de la probabilité a posteriori* comme critère d'alignement global. Trois hypothèses sont ensuite utilisées pour développer ce critère. La première hypothèse stipule l'indépendance locale de la distribution de probabilité $p(Y|a)$. La deuxième lie la distribution de la probabilité conditionnelle $p(Y_{\delta s}|a)$ au modèle transformé $\xi_s(M,a)$ au site s par la fonction $p(Y_{\delta s}|a) \approx \phi_s(Y_{\delta s}, \xi_s(M,a))$ où δs désigne un système de voisinage centré sur s . La troisième suppose l'indépendance mutuelle des $Y_{t \in \delta s}$ sachant le modèle transformé $\xi_s(M,a)$ au site s . La forme finale du critère est distribuée et redondante. L'interaction locale d'un site du modèle avec ses voisins dans l'image intègre deux intéressantes mesures de proximité et de similarité de l'orientation utilisées dans plusieurs techniques de regroupement basées sur l'organisation perceptuelle (voir annexe B).

Dans une deuxième étape et en utilisant une analogie avec la mécanique dynamique, on traduit la distribution d'interactions par un champ de force s'appliquant sur le modèle afin de le transformer. L'ensemble des forces locales développe une résultante de translation, un couple de rotation et une distribution de tension appliquée au modèle dans toutes les directions. La solution est trouvée par l'intégration temporelle de trois équations dynamiques représentant respectivement le mouvement de translation et de rotation du modèle et la mise à l'échelle dans la direction subissant la tension maximale. La mise à l'échelle dans toutes les directions cause implicitement une transformation de cisaillement.

Les transformations de rotation et de mise à l'échelle sont retardées jusqu'à ce que la force de translation soit inférieure à un seuil prédéterminé. La mise à l'échelle est aussi retardée par rapport au couple de rotation. Cette procédure hiérarchique de recherche d'équilibre mécanique améliore la capacité de convergence vers une solution globale.

La technique fonctionne avec différents types de représentation de modèle (de frontières, d'intensité) et utilise directement le gradient de l'image sans nécessiter une étape de pré-extraction de primitives.

La technique est expérimentée avec succès en utilisant des modèles de frontières et des modèles d'intensité à aligner avec des images synthétiques bruitées, des images de rayon-X et des images photographiques. La technique s'est montrée stable et robuste par rapport à différents types d'images présentant de grandes disparités et beaucoup de bruit.

Dynamic Model for Global Alignment and Scaling

Hail Mallouche, Jacques A. de Guise†, Yves Goussard

Biomedical Engineering Institute, Ecole Polytechnique of Montreal

P.O.Box 6079, Station "Centre-Ville," Montreal, Quebec, Canada, H3C 3A7

†Automated Production Engineering, Ecole de Technologie Superieure

1100, Notre-Dame West, Montreal, Quebec, Canada, H3C 1K3

Correspondence should be sent to:

Hail Mallouche

Biomedical Engineering Institute

Ecole Polytechnique of Montreal

P.O.Box 6079, Station "Centre-Ville"

Montreal, Quebec, Canada

H3C 3A7

Phone: (514) 340-4198

Fax: (514) 340-4611

Email: hail@grbb.polymtl.ca

Dynamic Model for Global Alignment and Scaling

Hail Mallouche, Jacques A. de Guise, Yves Goussard

Abstract. Global alignment techniques are either direct or iterative, and can generally be feature-based or intensity-based. Feature-based techniques are usually preceded by an extraction-grouping process that largely limits their utility. Many of these techniques are based on feature matching that makes the alignment sensitive to faulty matches. Moreover, intensity-based techniques are sensitive to image formation assumptions and noise. Finally, these techniques encounter many difficulties when dealing with large disparity situations.

In this paper, we present a dynamic model for global alignment and scaling. The alignment is done by optimizing a global *a posteriori* criterion developed without any particular assumption about the image formation process. The criterion has a distributive form measuring the local similarity (proximity and orientation) between the observed image and the model. The resulting potential field generates a force field that transforms the model. This field engenders a translation force, a rotation moment, and a stress distribution for scaling and skewing. These three forces are then time integrated using a second order mechanical system. Forces are also time scheduled in a hierarchical scheme to help preventing the search from being trapped in local optima. The approach was successfully tested for aligning 2-D wire frame models and intensity images. Unlike other techniques, our dynamic model is very robust and stable with respect to noisy images and large disparity conditions, and works well even with complex images. Finally, the approach can be used as feature-based or intensity-based,

and can be easily upgraded to 3-D applications.

Index Terms. Global and dynamic alignment, probabilistic model, affine transformation, feature-based and intensity-based methods, hierarchical alignment.

3.1 Introduction

Global alignment is of great interest in search methods for inference and matching problems in computer vision [1,2,4,5]. It consists of applying linear or non linear global transformations to a scene model in order to align it with the observed data. Pre-alignment step is helpful for many non convex search spaces in computer vision problems, and particularly when using descent search methods [6,10,11]. Accordingly, the global alignment allows us to position the initial solution as close as possible to the hill containing the optimal solution. This means that alignment methods can give a rough estimate for any subsequent matching process. However, despite the reduced dimension of their search space, many global alignment processes can equally be trapped in local optima. In addition, the sensitivity of the alignment process has to be considered for the stability of the search process. This stability can be compromised depending on the nature of the data and model representation and their mutual disparity.

Many global alignment methods have been proposed for positioning, recognition, tracking, motion estimating and registration [4,5,7,8,12,13,20], or as a first level process before matching [1,11]. These methods can be direct or iterative [2,5]. Direct methods are

based on invariant properties of particular transformations as for example the moments of inertia and the Fourier descriptors [5,7,15,20]. On the other hand, iterative methods seek to optimize a global criterion measuring the consistency between the model and the transformed image with respect to the alignment parameters [4,8,13,16]. These methods can be intensity-based or feature-based. The former methods are sensitive to intensity background variations, illumination conditions, and many other parameters of the image formation process [1,19]. These methods also encounter serious problems when dealing with large disparity situations and noisy data. On the other hand, feature-based methods generally need to be preceded by a good boundary extraction technique and even a grouping technique unless the model and the observed data are already given in boundary representations. This makes the technique sensitive to many parameters as, for example, the boundary fragmentation, the feature matching error and the noise. Additionally, some techniques use a point-to-point correspondence that makes them sensitive to occlusion or spurious data [4,8].

In this paper, we present a new global alignment technique. Our approach is iterative and uses a *maximum a posteriori* (MAP) estimator. This technique can be applied to any combination of boundary or intensity representations for the model and for the observed data, and works in 2-D and 3-D alignment problems. In addition, the proposed technique is robust and stable with respect to noise and works successfully with large disparity problems.

The transformed image is submitted to a distributed force field computed to measure the perceptual likeness between the state solution and the observed data. This means that the solution is continuously attracted by the force field until reaching the equilibrium. Submitted to this force field, the state solution undergoes four different linear transformations: (i)

translation, (ii) rotation, (iii) scaling in two directions, and (iv) skewing. The transformations are organized hierarchically to present higher stability and guarantee the convergence to the global solution. The solution of the global alignment is given when the second order dynamic system reaches its equilibrium.

In the next section, we present an overview of some important alignment techniques. The assumptions and the foundations of these techniques are presented and their limitations are discussed. Our approach is exposed in Section 3.3. This includes the MAP criterion, the dynamic model, and the time integration technique with its controlled version and intensity-based version. The implementation and the experimental results with synthetic and real images are presented in Section 3.4. Finally, the results and the approach are both discussed in Section 3.5.

3.2 Global alignment methods

In this section, we present an overview of the alignment methods related to our approach, and we discuss their limitations and the common points with our technique.

Alignment techniques based upon the adjustment of the global moments use moment invariant properties [5,7,15]. Let $Y(x,y)$ be a 2-D function and $\xi(Y,a)$ denote a \mathbb{R}^2 to \mathbb{R}^2 transformation where a is the parameter vector of the transformation. Let $M_{pq}(Y)$ be the conventional definition of the 2-D moment of order (p,q) for the 2-D function Y .

Let M_{pq}^ξ denotes the (p,q) moment of the transformed function $\xi(Y,a)$. The main idea of

the moment invariant techniques consists of expressing the moment M_{pq}^ξ as some nonlinear combination of the moment set $\{M_{rs}; r, s=0, 1, 2, \dots\}$ of the original function Y . For example, consider the scaling transformation $\xi(Y, a) = Y(x_1/a_1, x_2/a_2)$, M_{pq}^ξ can be written as:

$$M_{pq}^\xi = a_1^{1+p} a_2^{1+q} M_{pq} \quad (3.1)$$

Here, the alignment method consists of computing several values of $\{M_{pq}\}$ and $\{M_{pq}^\xi\}$, and then solving the invariant relations for the transformation parameters, e.g., scaling relation (3.1). This technique is particularly sensitive to the mass distribution in the image and therefore acts badly when applied to noisy images. Moreover, the alignment by rotation assumes that the viewed object has an elongated geometrical shape. Consequently, these methods lack precision when dealing with non elongated objects.

The alignment using Fourier descriptors is based on matching the model and the image descriptions [15,16,20]. Here the features to align are the silhouettes C^M and C of the model and the observed object, respectively. Both silhouettes are described by curvilinear complex functions $z^M(s)$ and $z(s)$ where s is the curvilinear scanning parameter normalized to 2π . We note that C^M and C have to be scanned using the same constant curvilinear speed. Now, each curvilinear curve can be decomposed into a sum of projections on an orthogonal set of functions. For the discrete Fourier descriptors (DFD), the set of orthogonal function is $\{e^{jns}, n=-\infty, \dots, 0, \dots, +\infty\}$. Let $\{A^M(n), n=\dots\}$ and $\{A(n), n=\dots\}$ denote the sets of Fourier

descriptors of C^M and C , respectively. Usually the series is truncated to the N lowest frequency coefficients. The alignment makes direct use of Fourier transform properties. The translation is done by adjusting the DC components. The scale is given by the ratio of the first harmonic components $\frac{|A(1)|}{|A_M(1)|}$. Finally, the in-phase rotation θ and the starting origin shift s_0

are aligned by the phases differences $e^{j\theta}$ and $e^{jk s_0}$ of the two largest coefficients of the Fourier descriptors, respectively. The DFD technique can also be used to match deformable silhouettes to the observed image [16]. Here, the alignment process is no longer direct, but instead becomes an iterative search in the parameter space. The DFD techniques are particularly sensitive to perturbations in the object boundary. In addition, the silhouette-based alignment requires a first level of boundary and silhouette extraction that can be troublesome when applied before global alignment and with noisy data.

Iterative alignment by minimizing the least squares error has been used to infer translation and rotation transformations or affine transformation so as to match a model to the boundaries extracted from the image [4,8]. These alignment techniques are feature-based which is a major drawback when this technique is to be used as the first step of a matching problem. In the former case, the mean quadratic error between the 2-D projection of a 3-D boundary model and the 2-D extracted boundaries is used to manipulate the pose parameters and minimize the error [8].

In the latter case, 3-D surfaces are aligned using affine transformations [4]. The technique consists of minimizing the least squares matching error between two discrete surface

functions. This requires a robust process of point-to-point correspondence and a good initial solution. This makes the technique sensitive to the point-to-point mismatch.

The alignment technique by maximization of the mutual information criterion is intensity-based [19]. The mutual information (MI) between the model Y^M and the transformed image $\xi(Y, \alpha)$ is given by:

$$I(Y^M, \xi(Y, \alpha)) = H(Y^M) + H(\xi(Y, \alpha)) - H(Y^M, \xi(Y, \alpha)) \quad (3.2)$$

H is used to denote marginal entropy and joint entropy as well. [9] uses a gradient descent method to maximize the MI with respect to alignment parameter α . However the first term of the RHS of the MI is independent from α which reduces the problem to the maximization of the sum of the second and third terms. Maximizing the former term $H(\xi(Y, \alpha))$ prevents accidental poses by smoothing the sampler distribution. Minimizing the latter term $H(Y^M, \xi(Y, \alpha))$ reduces the error between the model and the transformed image. This technique has two major limitations [19]. (i) It is very sensitive to the conditions of illumination and to the image formation process due to its intensity-based nature. (ii) The alignment works well only when the initial hypothetical pose is close to the true pose. This drawback is due the capture range of alignment that is not large enough to cope with any randomly chosen initial state and thus to ensure convergence to the true solution. Therefore, an additional rough estimator is required to propose some adequate initial pose.

Elastic models combined with optical flow techniques have been also used for motion recovery [12,13,18]. This technique uses an elastic finite element model. The approach is interesting for combining both rigid and non rigid motion dynamics together. The external

forces applied to the model are calculated using the optical flow from a sequence of video images. Then, the dynamic equation of the system is transformed in order to increase the computational efficiency. For this purpose, the stiffness matrix is decomposed into its eigenvector/eigenvalue representation known also as free vibration mode or frequency mode transformation. The lowest frequency modes are always the rigid-body modes that correspond to translation and rotation movements. The next higher frequency modes are the first smooth deformations of the whole-body that leave the center of mass and rotation invariant. The high frequency modes are discarded to reduce dimensionality. Two kinds of time integrating equations have been tested: the kinematic and the dynamic differential equations. The major limitations of this technique are the sensitivity to the optical flow estimate, to the initial state, and additionally the high computational cost of finite element modeling.

3.3 Method

In this paper we propose a global alignment technique for 2-D models with 2-D intensity images. The technique can be straightforwardly extended to 3-D alignment problems. Our goal is to develop an iterative technique having the following properties: (i) The alignment is realized by transforming the model to match the observed image and can be applicable to rigid, scalable or deformable models; We choose a linear transformation that has six degrees of freedom in the 2-D image plane. So, the model can undergo four different 2-D transformations: translation, rotation, scaling in two non orthogonal axes and skewing. (ii)

The technique should be robust and depend not on pre-processing of the data as boundary extraction and feature grouping; Therefore, we choose to directly match the model with the continuous gradient field of the Gaussian filtered image. (iii) The technique has to be based on the optimization of one global objective function; We choose a stochastic framework to build the global criterion. (iv) Finally, we wish to use a descending search method offering a high assurance of convergence even when the start state is far from the true solution; For this purpose, we choose a mechanical analogy to translate the model-to-image interaction where the dynamic manipulation of the system is described by second order differential equations making the system behave as a second order gradient method. The mechanical analogy helps to understand how external forces act to align and stress the model. This can be attractive for many applications as tracking moving and articulated objects. In addition, the likelihood of convergence to global solutions is improved by using a hierarchical scheme that consists of applying the permitted transformations in a time scheduled manner and by using a long range interaction neighborhood.

The technique uses a wire frame model to match directly the intensity image [10,11,17,18]. The model is made up of a set of Jordan closed curves. Each curve represents an intensity discontinuity of the model projection in the image plane. This wire frame representation is a basic aspect of our approach. Such a representation is helpful when estimating the image formation process is difficult. Indeed, in early vision, some parameters such as the illumination conditions are mostly unavailable. In addition, many other parameters of image formation cannot be easily estimated, e.g., surface reflective type when considering opaque objects, or attenuation and diffraction densities for semitransparent scenes.

In Section 3.3.1, we develop the probabilistic criterion for global alignment. In Section 3.3.2, we translate the criterion as a second order dynamic model. The resulting forces are presented in Section 3.3.3. Sections 3.3.4 and 3.3.5 are devoted to the time integration and the hierarchical scheme, respectively. In Section 3.3.6, we present four extensions of our method.

3.3.1 The stochastic model

In this section, we develop a probabilistic criterion for global alignment using a *maximum a posteriori* (MAP) estimator. MAP criterion has been extensively used in computer vision. It is attractive because it can help formulate many visual problems as a global search process. Let Y be the observed image and M be the *a priori* wire frame model. The deterministic image formation process is described by the function $\xi(M, \alpha)$ where α denotes a possibly random parameter vector made up of the global geometric alignment (pose) and transformations (scaling and skewing). In order to account for a stochastic component in the image formation process, we assume that the conditional probability distribution $p(Y|\alpha)$ is known and is a function $\phi(Y, \xi(M, \alpha))$. The exact form of this distribution depends on the distribution of the corrupting noise and on the nature of the corrupting process (e.g., additive, multiplicative, etc.). The problem consists of finding the parameter vector α such that the adjusted model $\xi(M, \alpha)$ best fits the observed image Y in some sense. The MAP criterion is given by:

$$\alpha^* = \arg \max_{\alpha} \{p(\alpha | Y)\} \quad (3.3)$$

Using the Bayes rule in (3.3) and neglecting the normalizing factor $p(Y)$ yields:

$$a^* = \arg \max_a \{ p(Y|a) \cdot p(a) \} \quad (3.4)$$

We now concentrate on the form of the conditional probability $p(Y|a)$. Let s denote any given site and δs a neighborhood of s . We introduce the following assumptions: (i) a form of neighborhood independence that allows us to write $p(Y|a) \propto \prod_s p(Y_{\delta s}|a)$; (ii) the possibility to express the above conditional probability as $p(Y_{\delta s}|a) \approx \phi_s(Y_{\delta s}, \xi_s(M, a))$; the above assumption essentially means that the behavior of the observed scene on δs is mainly determined by the value of the transformed model at site s ; (iii) the possibility to further factor the above expression as $\phi_s(Y_{\delta s}, \xi_s(M, a)) = \prod_{t \in \delta s} \psi_{s,t}(Y_t, \xi_s(M, a))$, which amounts to assuming the independence of the pixels of δs given the image model at site s . Introducing these assumptions into the MAP criterion yields:

$$a^* = \arg \max_a \left\{ \prod_s \prod_{t \in \delta s} \psi_{s,t}(Y_t, \xi_s(M, a)) \right\} \quad (3.5)$$

Let us now assume that the observed data is not exactly the intensity image Y but instead a directional function $\vec{b}(Y)$ of Y , e.g., the directional gradient of the intensity image.

Therefore, here we replace Y_t and $\xi_s(M, a)$ with the local directional gradient vector $\vec{b}(Y_t)$

and the curve direction $\vec{b}(\xi_s(M, a))$, respectively. Finally, we take the logarithm of the

functional to write the criterion as:

$$\alpha^* = \arg \max_{\alpha} \left\{ \sum_s \sum_{t \in \delta s} \ln \psi_{s,t} \left(\vec{b}(Y_t), \vec{b}(\xi_s(M, \alpha)) \right) \right\} \quad (3.6)$$

The functional of the criterion (3.6) is the objective function and is denoted E . The inner term is denoted $W_{s,t}(Y_t, \xi_s(M, \alpha)) = \ln \psi_{s,t}(\vec{b}(Y_t), \vec{b}(\xi_s(M, \alpha)))$. For the explicit form of $W_{s,t}$,

we write $W_{s,t} = V_{s,t} - \ln Z$ where $Z = \sum_{Y_t, \xi_s(M, \alpha)} e^{V_{s,t}}$ is a normalizing constant. Then, the objective

function becomes $E = \sum_{s \in L} \sum_{t \in \delta s} V_{s,t}$. $V_{s,t}$ can be interpreted as a local similarity measurement

between the model and the image and referred to as site-to-site potential difference between the model and the image. Perceptual organization uses geometric constraints for feature grouping [2,5]. Proximity and orientation can be used to measure the similarity or likeness between the vectors $\vec{b}(Y_t)$ and $\vec{b}(\xi_s(M, \alpha))$. For this purpose, we propose the following expression for $V_{s,t}$:

$$V_{s,t} = e^{-\frac{|s-t|^2}{\sigma^2}} \left(\vec{b}(\xi_s(M, \alpha)) \cdot \vec{b}(Y_t) \right) \quad (3.7)$$

The exponential term is the proximity measurement between the site s of the model and the site t of the image where σ^2 is the variance of the proximity function. The scalar product between $\vec{b}(Y_t)$ and $\vec{b}(\xi_s(M, \alpha))$ measures the similarity of the local orientation between the image and the model. The estimate of α is obtained through maximization of E . The search space dimension is equal to the order of the allowable transformation. This global alignment

assumes knowing the object identity with which to match. The global alignment can be seen as a sort of reduction in the search space dimension, or as a parametric search in a coarse space. In the next section, we discuss different search techniques and present our dynamic search model.

3.3.2 The dynamic model

Different optimization methods can be used to search for the global solution of E . Search methods can be stochastic or deterministic. A stochastic search can overcome many local optima and reach acceptable global optimum [6]. However, these methods are computationally burdensome. Deterministic search techniques are generally less time consuming but can be trapped in local optima. The choice of a technique mainly depends on the following elements: (i) the state space dimension and complexity, (ii) the transition process, and (iii) the start solution. The transition process affects the search space convexity. A dense transition process yields a strong convexity. When the initial solution belongs to the global hill, deterministic methods will necessarily converge. Finally, a large state space dimension usually means that the search space is complex with many local optima. All these factors have to be considered when choosing the search method.

Global alignment has the following advantages. The search space dimension is reduced to the order of the alignment process. The transformations included in the model are composed of translation, rotation, scaling and the skewing. For a 3-D alignment, this is equivalent to 12 degrees of freedom that will be reduced to six when considering a 2-D alignment problem. On the other hand, the local neighborhood δs will affect the transition

process. Larger δs means long range interaction between the model and the observed image and consequently a dense transition process. For these reasons, global alignment methods generally use deterministic search methods that have an affordable computational cost. Note that each deterministic method has its own specific behavior. In fact, even in global alignment problems, deterministic methods can be trapped in local optima.

In this section, we propose a deterministic dynamic search technique based on analogy with mechanics to optimize the global alignment OF. This technique uses a second order dynamic modeling combined with delayed transformations to enhance the convergence to global solution. We now present the mechanical model of the dynamic technique.

Let m and J be the mass and the moment of inertia of the model (wire frame, set of discontinuous lines or intensity model), respectively. m and J can be calculated either from the surface representation or the perimeter representation of the wire frame model. We now recall the OF E . From a mechanical standpoint, the potential difference $V_{s,t}$ can be interpreted as an attraction force between sites s and t . Our mechanical model is then based on a distributed field force $F = \{\vec{F}_s, s \in \xi(M, a)\}$ substituted for the OF E . The local force \vec{F}_s applied to site s has also a distributive form $\vec{F}_s = \sum_{t \in \delta s} \vec{F}_{s,t}$, where:

$$\vec{F}_{s,t} = e^{-\frac{1s-tl^2}{\sigma}} \left(\vec{b}(\xi_s(M, a)) \cdot \vec{b}(Y_t) \right) \vec{u}_{st} \quad (3.8)$$

The dot sign and $\vec{u}_{st} = \frac{s-t}{\|s-t\|}$ denote the inner product and the unit vector defined by the

direction from s to t (*Figure 3.1*), respectively. The model will then undergo the permissible transformations until reaching an equilibrium point. The difference between $V_{s,t}$ and the local force $\vec{F}_{s,t}$ is the unit vector \vec{u}_{st} . If the local mass is assumed fixed to same value on each site of the model, $\vec{F}_{s,t}$ can be seen as a natural expression of the force field corresponding to $V_{s,t}$ or as a duality between force and potential difference. In order to test intuitively the stability of the equilibrium that can be reached, we consider a local gradient field hill $\vec{b}(Y_{\delta s'})$ where s' is the site identifying the maximum of the hill, and a local wire frame segment $\vec{b}(\xi_{\delta s})$ (*Figure 3.2*). If $\vec{b}(\xi_{\delta s})$ is aligned to $\vec{b}(Y_{\delta s'})$ and s coincides with s' then V_s is the optimum in the local hill (equilibrium position $\vec{F}_s=0$). If we slightly disturb $\vec{b}(\xi_{\delta s})$ from the equilibrium position s' , the resultant force \vec{F}_s will be oriented toward s' which will reposition s to s' .

Let l , $\bar{\theta}$ and \vec{g}_α be the translation vector, the rotation angle around the gravity center s_{GC} of the model and the scaling factor undergone by the model, respectively. Here, \vec{g}_α is the scaling vector in direction α and centered on s_{GC} . The deterministic movement of the model is then describe with a set of second order dynamic equations [13, 14]. On one hand, the translation and rotation equations are given by the following:

$$\begin{aligned}\vec{l} &= \frac{\vec{F}}{m} - f \vec{l} \\ \vec{\ddot{\theta}} &= \frac{\vec{C}}{J} - f \vec{\ddot{\theta}}\end{aligned}\tag{3.9}$$

\vec{F} and \vec{C} are the total force and the total torque simultaneously applied to the model by the field of force F , respectively. f is the damping coefficient. On the other hand, we chose to combine the effect of scaling and skewing in one linear transformation G . G is a 2x2 matrix when alignment is in the image plane. Let $s = [x_{s,1} \ x_{s,2}]^T$ and $s' = [x_{s',1} \ x_{s',2}]^T$ be two points in the initial reference of the model before transformation and in the transformed reference, respectively. We write $s' = G \cdot s$. Let us consider a scaling operation centered on the origin with scaling factor g and direction α . Let R_α be the rotation matrix to the direction α and around the origin. $R_{-\alpha}$ denotes the inverse of R_α . The scaling g in the direction α can then be written as:

$$G_{\alpha,g} = R_{-\alpha} G_{s_1,g} R_\alpha \tag{3.10}$$

$G_{\alpha,g}$ and $G_{s_1,g}$ are the transformations of scaling in the direction α and direction s_1 . $G_{s_1,g}$ is a diagonal matrix with the first diagonal element equals g and the other diagonal elements are set to one. For the 2-D alignment, $G_{\alpha,g}$ will then be given by:

$$G_{\alpha,g} = \begin{bmatrix} 1+(g-1)\cos^2\alpha & (g-1)\cos\alpha\sin\alpha \\ (g-1)\cos\alpha\sin\alpha & 1+(g-1)\sin^2\alpha \end{bmatrix} \tag{3.11}$$

If the scaling operation is centered on a point different from the origin, like, for example, the

gravity center s_{GC} of the model, the transformation $s' = G_{\alpha,g} \cdot s$ is then replaced by the following:

$$s' = G_{\alpha,g} \cdot s - [G_{\alpha,g} - I] \cdot s_{GC} \quad (3.12)$$

The mechanical interpretation of the directional scaling is equivalent to stretching or compressing the model under the effect of antagonist forces. The resultant stress for direction α is denoted \vec{S}_α . The corresponding dynamic equation can be a second order differential one as for the rigid motion. An inner elastic tensor can also be incorporated into the equation. However, no elastic component is wanted in such a global alignment because it can create a strong bias toward the initial state of the wire frame model. We chose to allow scaling in any direction. However, successive arbitrary directions are not mutually orthogonal which means that scaling in one direction also affects other non orthogonal directions and additionally creates a skewing effect. In fact, we choose to combine the iterative scaling G_{α_i, g_i} , $i=0, \dots, n$, in one progressive linear transformation $G = \prod_{i=0, \dots, n} G_{\alpha_i, g_i}$ which makes the use of a second order differential equation without damping component convenient to describe the dynamic behavior of the stress. Therefore, we write:

$$\ddot{g}_i = \frac{|\vec{S}_{\alpha_i}|}{m} \quad (3.13)$$

The difference with the dynamic system for the translation and rotation is that when scaling, for example, in α_i direction, the first derivative \dot{g}_{i-1} of the iteration $i-1$ is always assumed

to be equal to zero.

3.3.3 The resultant forces

Three kinds of attraction forces result from the already defined field of attraction \vec{F}_s . These forces are computed to manipulate the three allowable dynamic transformations. The corresponding forces of translation, rotation and scaling, are \vec{F} , \vec{C} and \vec{S}_α , respectively. In this section, we see how to deduce these forces from the field \vec{F}_s .

We recall that the field of attraction \vec{F}_s is defined over each site s belonging to the boundary set of the model. Consequently, the translation movement results from the sum \vec{F} of \vec{F}_s over the sites of the model. This yields:

$$\vec{F} = \sum_s \vec{F}_s \quad (3.14)$$

In addition, both dynamic equations (3.9) of translation and rotation contain energy dissipation (damping) components that are proportional to the global first derivatives $\vec{\dot{l}}$ and $\vec{\dot{\theta}}$, respectively. The damping coefficient f is adjusted to give the dynamic model an optimal

behavior in the sense of response time and stability of convergence. The total torque \vec{C}

applied to the model is given by:

$$\vec{C} = \sum_s \vec{F}_s \times (s - s_{GC}) \quad (3.15)$$

where $s - s_{GC}$ is the distance vector between the point s and the gravity center of the model s_{GC} , and \times denotes the cross product. Finally, the stress force \vec{S}_α centered on s_{GC} in the direction α is given as a difference between two antagonistic forces that both are parallel to the direction α . To compute these forces, we split the space into two disjoint half spaces P_α^+ and P_α^- using the plane defined by its normal direction α and passing through the point s_{GC} (see *Figure 3.3*). For 2-D alignment, spaces P_α^+ and P_α^- represent two half-planes, and the splitting plane is reduced to a splitting line. \vec{S}_α is then given by:

$$\vec{S}_\alpha = \left| \left(\sum_{s \in P_\alpha^+} \vec{F}_s - \sum_{s \in P_\alpha^-} \vec{F}_s \right) \cdot \vec{u}_\alpha \right| \vec{u}_\alpha \quad (3.16)$$

\vec{u}_α is the unit vector defining the direction α .

3.3.4 Iterative time integration

The model submitted to the time varying attraction field \vec{F}_s , and its resulting global force \vec{F} , torque \vec{C} , and stress \vec{S}_α , is iteratively transformed until reaching an equilibrium state. Our

search method is an ordinary discrete version of the time integration of the second order dynamic equations [12,13]. The roughest method uses a truncated Taylor expansion that yields:

$$\begin{aligned}\bar{l}(t) &= \left(\frac{\bar{F}(t)}{m} - \bar{l}(t-1) \right) \frac{dt^2}{2} - \bar{l}(t-1) dt + \bar{l}(t-1) \\ \bar{\theta}(t) &= \left(\frac{\bar{C}(t)}{J} - \bar{\theta}(t-1) \right) \frac{dt^2}{2} - \bar{\theta}(t-1) dt + \bar{\theta}(t-1) \\ g_i(t) &= \frac{|\bar{S}_{\alpha_i}(t)|}{m} \frac{dt^2}{2}\end{aligned}\tag{3.17}$$

where t denotes the current time or iteration and $t-1$ is the previous iteration. This is a second order integration technique. Higher order techniques, e.g., Runge-Kutta, can also be used. These techniques can improve the precision of the trajectories without any significant effect on the convergence speed. In fact, the inaccuracy of the integration can be considered as a weak stochastic perturbation component in the search process. For $t=0$, the initial speeds of the translation and rotation are set to zero.

The integration of the scaling dynamic equation is slightly different from the former ones. At each time step t we choose the scaling direction α_t from the N_S discrete directions. Then we stochastically sample from the distribution that is proportional to the computed stress in the N_S different directions. The technique is simple and increases the likelihood of sampling in the directions of high stress, and introduces a stochastic component in the search method. In addition, at each integration step, the term $\dot{g}(t-1)$ is set to zero which reduces the time expression of the scaling dynamic to a simple first order form.

3.3.5 Hierarchical transformations

The hierarchical structure aims at developing a process of concatenated transformations to reduce the possibility to be trapped in local optima during the solution search. The major advantage of transforming the global alignment problem into a force equilibrium search is that it permits to explore the search space using the principal components of the movement. In addition, we reduced the scale-skew transformation to search only in one scaling direction at each time-step.

The scheme can be seen as a mechanical object moving under the resultant forces of an external field. So from a mechanical standpoint, reaching the global optimum is more likely when first aligning the gravity centers. When some acceptable equilibrium is reached, the allowed set of transformation is then expended to include the rotation. Again, when certain level of equilibrium is reached, the scaling and skewing transformations are added to the set of transformation. This technique of equilibrium delayed transformations is natural from a mechanical standpoint. It can be also seen as some global to local search method that can improve global convergence for a wide range of starting solutions.

So, we can see the problem in an imbedded scheme of geometric movements. This means we need to organize the transformations with respect to which one can offer a higher assurance of converging to global equilibrium when applied in a hierarchical scheme. Here, it seems much better at first to apply the translation alone until the global resultant force $|\vec{F}|$ decreases below a fixed threshold. During the decrease of $|\vec{F}|$, the rotation moment \vec{C} can gradually be applied until reaching its full strength. This thresholded (or controlled) action can

be similarly repeated with the stress force \vec{S}_α .

Algorithmically speaking, the controlled transformations can be written as follows. Let th_F and th_C be the thresholds associated with the force \vec{F} and the moment \vec{C} , respectively. The new controlled force, moment and directional stress, \vec{F}^N , \vec{C}^N and \vec{S}_α^N are given by:

$$\vec{F}^N = \vec{F} \quad (3.18)$$

$$\vec{C}^N = \begin{cases} 0; & \frac{|\vec{F}|}{m} > th_F \\ \vec{C} \cdot \left(1 - \frac{|\vec{F}|}{m th_F} \right); & \frac{|\vec{F}|}{m} < th_F \end{cases} \quad (3.19)$$

$$\vec{S}_\alpha^N = \begin{cases} 0; & \frac{|\vec{F}|}{m} > th_F \text{ or } \frac{|\vec{C}|}{J} > th_C \\ \vec{C} \cdot \left(1 - \frac{|\vec{F}|}{m th_F} \right) \cdot \left(1 - \frac{|\vec{C}|}{J th_C} \right); & \frac{|\vec{F}|}{m} < th_F \text{ and } \frac{|\vec{C}|}{J} < th_C \end{cases} \quad (3.20)$$

Here, the force \vec{F}^N is not constrained. Thus, the translation is continuously allowed during the dynamic alignment. The moment \vec{C}^N is controlled by the force \vec{F} , which is equivalent to constraining the rotation movement by the translation. The rotation movement is allowed when the translation force module $|\vec{F}|$ goes below threshold th_F . Finally, the stress \vec{S}_α^N is controlled by both \vec{F} and \vec{C} , for limiting the scale transformation while \vec{F} and \vec{C} have not gone below the force and moment thresholds, respectively. When $\vec{b}(\xi(M, a))$ and $\vec{b}(Y)$ are

normalized to one, the values of th_F and th_C can be set to some small fractions of estimated upper bounds for $|\vec{F}|$ and $|\vec{C}|$, respectively.

3.3.6 Extensions

Wire frame model

We recall the expression of $V_{s,t}$ where the term $\vec{b}(\xi_s(M, a))$ represents the boundary vector at the site s of the projected model $\xi(M, a)$. In fact, determining the image formation function ξ is usually difficult or even unpractical. Instead, we simply replaced $\vec{b}(\xi_s(M, a))$ at the site s by the unit vector $\vec{u}_{a_s(M)}$ of the boundary direction of the transformed wire frame model $a_s(M)$. This means that a simple wire frame (boundary) representation of the model is sufficient to allow the alignment technique to work well.

Articulated objects

Our approach can be easily extended for alignment of articulated objects [2, 13]. Here we assume that the model is assembled using a set of subobjects. The assemblage is described with a 3-D relational graph [2]. The fundamental points of this extension are the followings.

- (i) The model is aligned rigidly to the image using the dynamic process developed in the previous sections. This means that all subobjects are glued together and undergo the same transformations throughout the dynamic process.
- (ii) After the equilibrium is reached, the subobjects are free to seek each their own alignment using the same dynamic process.
- (iii)

However, the subobjects have to respect the relational constraints of the original model during their alignments. At this level, the approach can incorporate mutual interactions that can represent the geometric, topologic or mechanical properties of the model's articulations.

Data to data alignment

Our method can be slightly modified in order to align two observed images, e.g., images issued from two different viewing positions of the same scene, or any pair of observed data.

This can also be useful for tracking moving objects in an image sequence. Let X be the image to be aligned to the reference image Y . For the image to image alignment version,

$\vec{b}(\xi_s(M, a))$, in both $V_{s,t}$ and $\vec{F}_{s,t}$, is simply replaced by the directional gradient vector

$$\vec{b}(X_s).$$

3-D alignment extension

Our method can be directly extended to the 3-D global alignment case. Either boundary or intensity-based representations can be used. The neighborhood of attraction becomes a 3-D ball instead of 2-D disk. The resulting forces are then given by a 3-D translation force, a 3-D torque vector and 3-D distribution of the directional stress. The dynamic behavior of the mechanical model is described using 3-D second order equations with dumping components for both translation and rotation transformations, and a 1-D second order equation applied in the direction of maximum stress.

3.4 Implementation and experimental results

3.4.1 Implementation

Our approach was implemented for a 2-D alignment scheme using the C programming language. Two versions were developed. The first one aligns 2-D wire frame models with 2-D intensity images. The second version aligns 2-D images with 2-D images. The first version is limited to wire frame models represented with a set of multiple Jordan curves [10,11]. Each Jordan curve (continuous closed uncrossing curve) is implemented as a coded chain. Let a model C^M and a site $s \in C^M$. The maximum radius of attraction or the neighborhood dimension $|\delta s|$ is set to a value bounded by an estimate of the maximum displacement between the image and the model. To reduce the computational time gradually, $|\delta s|$ is relaxed throughout the iterative search process. This is usually acceptable because of the reduction of the disparity between the image and the transformed model $a(C^M)$. The total transformation applied to site s is written as $a(s \in C^M) = \vec{I} + (s_{GC} + R_{\theta} \cdot G \cdot (s - s_{GC}))$. Both rotation and scaling-skewing operations are centered on the model centroid s_{GC} . Each iteration is one time-step of the discrete time integration of the dynamic equations (3.17). The translation $\vec{I}(t=0)$ and rotation $\vec{\theta}(t=0)$ are set to zero, and the scaling matrix $G(t=0)$ is set identity. We note that both gradient fields $\vec{b}(Y)$ and $\vec{b}(C^M)$ (or $\vec{b}(X)$) are calculated before starting the alignment process. Throughout the iterative alignment, fields $\vec{b}(a(C^M))$ or $\vec{b}(a(X))$ are directly computed by $a(\vec{b}(C^M))$ or $a(\vec{b}(X))$, respectively.

Each iteration consists of the following steps:

- (i) Scan sites $s \in a(C^M)$ (or $a(X)$) to compute the force field $\vec{F}_s = \sum_{t \in \delta s} \vec{F}_{s,t}$.
- (ii) Compute the force \vec{F} , moment \vec{C} , and stress distribution \vec{S}_α , $\alpha \in \pi$, using equations (3.14), (3.15) and (3.16), respectively.
- (iii) Compute the resulting translation, rotation and scaling, using dynamic equations (3.17) combined with the hierarchical scheme (3.18), (3.19) and (3.20).
- (iv) Check termination criterion (either comparing if $|\vec{F}|$, $|\vec{C}|$ and $|\vec{S}_\alpha|$ reached determined thresholds, or using a preset number of iterations).

Let us now consider the computation time issue. For the wire frame-based version, this time is of the order of $O(K \cdot (m+n) \cdot |\delta s|)$ where K is the number of Jordan curves constituting the model and (m, n) are the 2-D dimensions of the model. The computational time is higher for the intensity-based version, and is of the order of $O(m \cdot n \cdot |\delta s|)$ because the image to align must be entirely scanned at each iteration to compute the new force field F . A significant reduction in the computation can be achieved by using a multiresolution data structure. Another avenue consists of substituting $\vec{b}_{\nabla^2=0}(Y)$ and $\vec{b}_{\nabla^2=0}(X)$, which are the respective products by zero-crossing map of the Laplacian-Gaussian convolution, for $\vec{b}(Y)$ and $\vec{b}(X)$. This sets to zero each site s where $\vec{b}(Y_s)$ or $\vec{b}(X_s)$ is not a local intensity discontinuity. Consequently, \vec{F}_s is straightforwardly set to zero if $\vec{b}_{\nabla^2=0}(X_s) = 0$. In the same manner, $\vec{F}_{s,t}$ is also set to zero if $\vec{b}_{\nabla^2=0}(Y_t) = 0$. The results of the accelerated implementation

were similar to the original version.

3.4.2 Results

Both wire frame-based and intensity-based versions were tested using synthetic and real images. To verify the robustness and stability of the approach with respect to the noise, we conducted several experiments with synthetic noisy images. *Figure 3.4* shows two alignment experiments of simple wire frames (rectangular forms). In both, we used noisy images with low contrast rectangular objects, while in addition the image of the second experiment presents a varying background. Panels *a* and *b* of *Figure 3.4* show the models superimposing the noisy images. The dynamic equations were integrated for 200 time-steps. Resulting aligned models are shown in panels *c* and *d* of *Figure 3.4*. The evolution of the transformation parameters during the alignment process of the first experiment of *Figure 3.4* is presented in *Figure 3.5*. We can see the delayed transformations in the order of translation, rotation then scaling and skewing. We note that alignment generated weak skewing deformations because final values of both G_{01} and G_{10} are smaller than 0.01. An excellent convergence was observed for all experiments.

The second set of experiments with wire frame models was performed with X-ray images of complex and non elongated vertebra phantom. Panels *a* and *b* of *Figure 3.6* depict a wire frame model superimposing two X-ray views. The model is hand drawn and represents the boundary projection of a free cylindric form (body) and a free planar form (process). The alignment results are presented in panels *c* and *d* of *Figure 3.6*. However, the result of *d* used

an articulated representation of the model and so dynamic movement was made up of two phases: glued and articulated. We note how the process has moved away from the vertebra body in d but stills articulated to the body to preserve the regional relational aspect of the initial model b . The relational constraint prohibited the alignment of the process to continue converging to an optimal state. Given the limitation mentioned above and even though the geometric shape and the starting position of the model were far from the real object, the convergence was consistent and the quality of the final solution was acceptable for both glued and articulated model.

We now present the results of the data-to-data alignment experiments. Panels a , b , c , and d of *Figure 3.7* respectively present a black binary club image, a deformed image of the club, the alignment result of b to a , and finally the difference between c and a . Here too, the integration time is 200 steps. Note that the alignment is almost perfect. *Figure 3.8* and *Figure 3.9* show two alignment experiments of gray level images with the same presentation as in *Figure 3.7*. In *Figure 3.8* the two images are two stereo views of a toy scene. The alignment compensates for -4.4 pixels and -18.0 pixels of displacement in the vertical and horizontal directions, respectively. In *Figure 3.9* the two images are the Lenna image and a deformed version of the same image. The alignment was accurate and the original image was precisely restored. The alignment parameters were 5x5 pixels of translation, 0.25 radians of rotation, 1.15 of scaling in vertical direction and 0.85 of scaling in horizontal direction.

All experiments were conducted using images of 128x128 pixels. Experiments were run on a sequential processor. However, in order to reduce the execution time and due to the distributive nature of the field of attraction, the iterative process can be straightforwardly

implemented on a parallel machine.

3.5 Conclusion

We showed that the proposed approach for global alignment offers many interesting properties. It can work well with a variety of image and model representations, and under difficult conditions. Our alignment technique was successfully tested with images of complex objects and varying contrast. Additionally, the robustness and stability while dealing with large disparity situations, and the distributive form of the criterion make this technique very attractive.

The line boundary representation is simple and can be easily extended to for 3-D surface alignment. The only information needed is the boundary orientation or instead the inner and outer regions of each curve or surface. The technique can be also used for open boundaries and even for fragmented boundaries on condition that these boundaries incorporate the direction information of the intensity discontinuities.

Our technique can also be used to align a reference image to an observed image where both are intensity level images. Alignment works directly on the gradient intensity fields of both the reference and the observed images. This makes the technique perfectly suitable to initial global alignment in many model-based recognition tasks. The technique is very helpful particularly when images are produced under different formation conditions or when the image formation process cannot be determined analytically.

Finally, our approach is directly related to [4] in the sense of the affine transformation. For

our approach, we also use the translation, rotation, scaling and skewing operations in a separate form, and not in one combined affine transformation. This allows us to decompose the problem into three distinct mechanical transformations. We argue that scheduled (or delayed) transformations can increase the probability of reaching a global optima when using deterministic search methods. Furthermore, we differ from affine technique by the representations of the image and the model. In their case, image and model are both 3-D surfaces. For us, the image can be any 2-D or 3-D intensity image, and the model can be any 2-D curve representation or a 3-D surface representation and even any intensity image. In addition, our approach uses second order dynamic equations to reach the global equilibrium that resembles the mechanical modeling used in elastic techniques [13], and does not need any process of point-to-point correspondence. This makes our approach simple, robust and directly useful for a wide range of applications.

Acknowledgments

HM was supported by the National Sciences and Engineering Research Council of Canada (NSERC), the "Fonds pour la formation de chercheurs et l'aide à la recherche" of Quebec, the Canadian Council of Professional Engineers-North American Life, and the cooperation France-Quebec, under scholarship numbers 167653, 961252, CCPE/NAL95-96, and 143690B, respectively. JADG was supported in part by the NSERC of Canada. YG was supported in part by the NSERC of Canada under Research Grant OGP0137417.

References

1. R. Bajscy and S. Kovacic, Multiresolution elastic matching, *Computer Vision, Graphics, and Image Processing* 46, 1989, 1-21.
2. D.H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, 1982..
3. H. Derin and H. Elliot, Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields, *IEEE Trans. Pattern Analysis and Machine Intelligence* 9(1), 1987, 39-55.
4. J. Feldmar and N. Ayache, Rigid, affine and locally affine registration of free-form surfaces, *International Journal of Computer Vision* 18(2), 1996, 99-119.
5. W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*,

- MIT Press, Cambridge, 1990.
6. B. Hajek, Cooling schedules for optimal annealing, *Mathematics of Operations Research* **13**(2), 1988, 311-329.
 7. M.-K. Hu, Visual pattern recognition by moment invariants, *IRE Transactions on Information Theory* **IT-8**, 1962, 179-187.
 8. S. Lavallée and R. Szeliski, Recovering the position and orientation of free-form objects from image contours using 3D distance maps, *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**(4), 1995, 378-390.
 9. D.G. Lowe, Three-dimensional object recognition from single two-dimensional images, *Artificial Intelligence* **31**, 1987, 355-395.
 10. H. Mallouche, J.A. de Guise and Y. Goussard, Probabilistic model of multiple dynamic curve matching for semitransparent scene of nonrigid objects, *submitted to the IEEE Trans. Pattern Analysis and Machine Intelligence*, 1996.
 11. H. Mallouche, J.A. de Guise and Y. Goussard, Probabilistic model of multiple dynamic curve matching for semitransparent scene, in *Proceedings, Vision Geometry IV Conference*, San Diego, SPIE 2573, 1995, 148-157.
 12. C. Nastar and N. Ayache, Frequency-Based Nonrigid Motion Analysis: Application to Four Dimensional Medical Images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **18**(11), 1996, 1067-1079.
 13. A. Pentland and B. Horowitz, Recovery of nonrigid motion and structure, *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**(7), 1991, 730-742.
 14. W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*. Cambridge University Press, 1988.
 15. A.P. Reeves, R.J. Prokop, S.E. Andrews and F.P. Kuhl, Three-dimensional shape analysis using moments and Fourier descriptors, *IEEE Trans. Pattern Analysis and Machine Intelligence* **10**(6), 1988, 937-943.
 16. L.H. Staib and J. S. Duncan Boundary finding with parametrically deformable models, *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(11), 1992, 1061-1075.
 17. D. Terzopoulos, Matching deformable models to images: direct and iterative solutions, in *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, DC, 1987, 164-167.
 18. D. Terzopoulos, J. Platt, A. Barr and K. Fleischer, Elastically deformable models, *Computer Graphics* **21**(4), 1987, 205-214.
 19. P.A. Viola, *Alignment by Maximization of Mutual Information*, MIT, AIL, A.I. Technical Report 1548, 1995.
 20. T.P. Wallace and P. Wintz, An efficient, three-dimensional aircraft recognition algorithm using normalized Fourier descriptors, *Computer Graphics and Image Processing* **3**, 1980, 99-126.

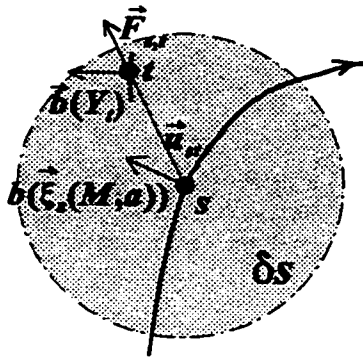


Figure 3.1. Attraction neighborhood δs centered on site s of the transformed model $\xi(M, a)$. The attraction force $\vec{F}_{s,t}$ between s and the boundary element $\vec{b}(Y_t)$ of the observed image Y at the site t is aligned with the unit vector $\vec{u}_{s,t}$.

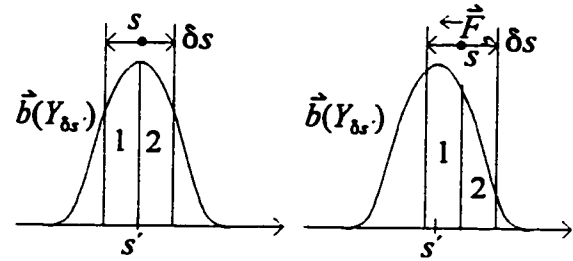


Figure 3.2. The dynamic model. $\vec{b}(Y_{\delta s})$ is 1-D slice of the observed local directional gradient. Left: surfaces 1 and 2 are equals, $\vec{F}_s = 0$ and the wire frame element s is in a equilibrium state over s' . Right: the equilibrium is disturbed and s is shifted from s' ; surface 1 is greater than surface 2, hence left component of \vec{F}_s is greater than the right one, and s will be attracted to s' .

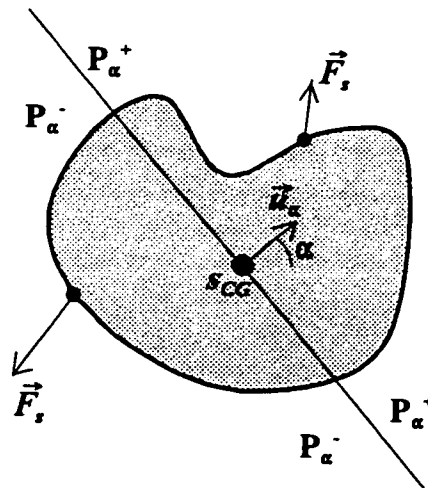


Figure 3.3. To compute the stress \vec{S}_α in the direction \vec{u}_α , the model plane is split into two half planes P_α^+ and P_α^- centered on the gravity center of the model.

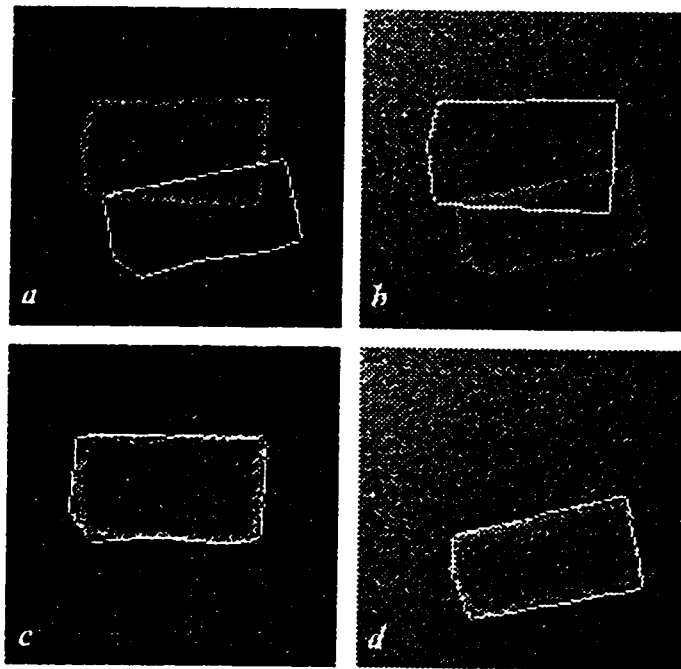


Figure 3.4. In the upper panels, the wire frame models are superimposed on two noisy synthetic images. The lower panels show the final solutions after 200 time steps superimposed on the corresponding noisy images.

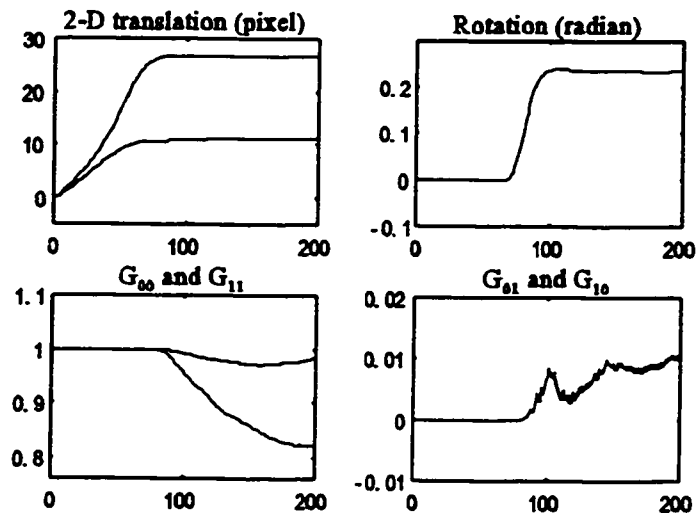


Figure 3.5. Evolution of the transformation parameters of the second alignment experiment of *Figure 3.1*. At the top left, top right, bottom left and bottom right we have the x and y translations, the rotation and the corresponding trend of the four terms of matrix G , respectively.

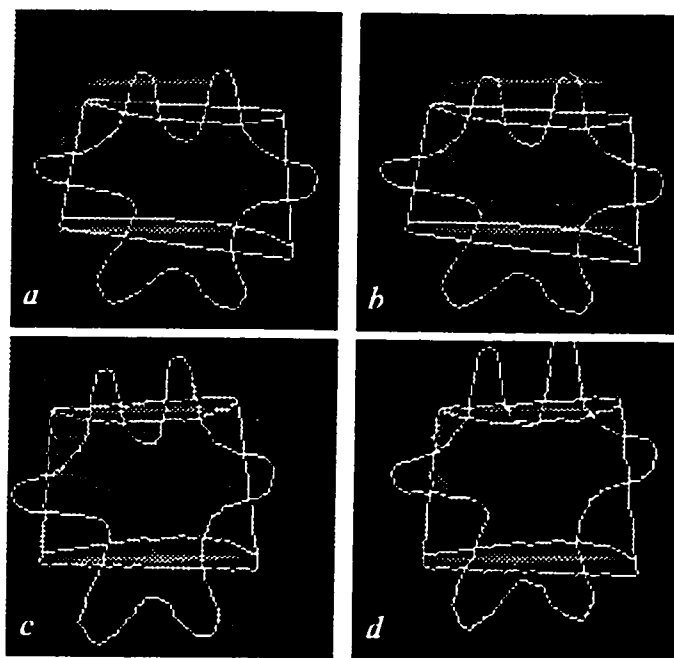


Figure 3.6. a and b: 2-D wire frame model superimposed over two X-ray images of a vertebra phantom taken from two slightly different poses. c and d: The corresponding alignment results after 200 time-steps. A 2-D articulated model is used for the result of d.

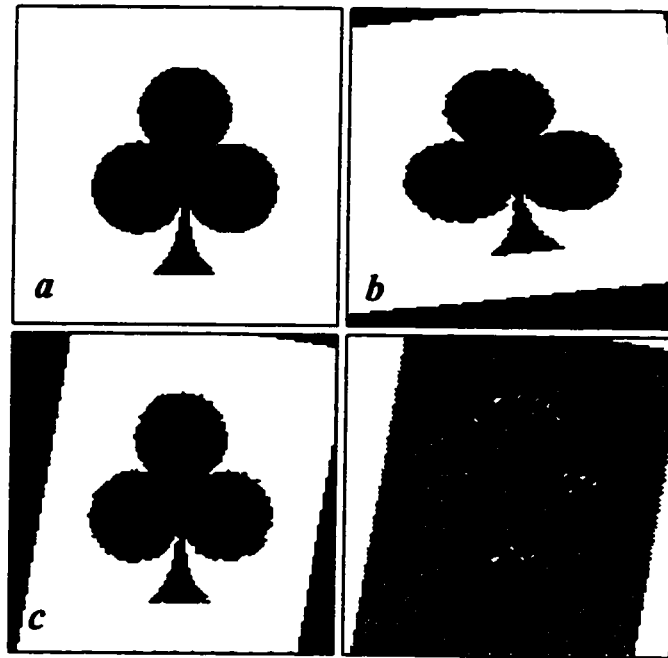


Figure 3.7. Binary image alignment. *a*: club image. *b*: the affine transformation of *a*. *c*: alignment result of *b* to *a* after 200 time-steps. *d*: difference between *a* and *c*.

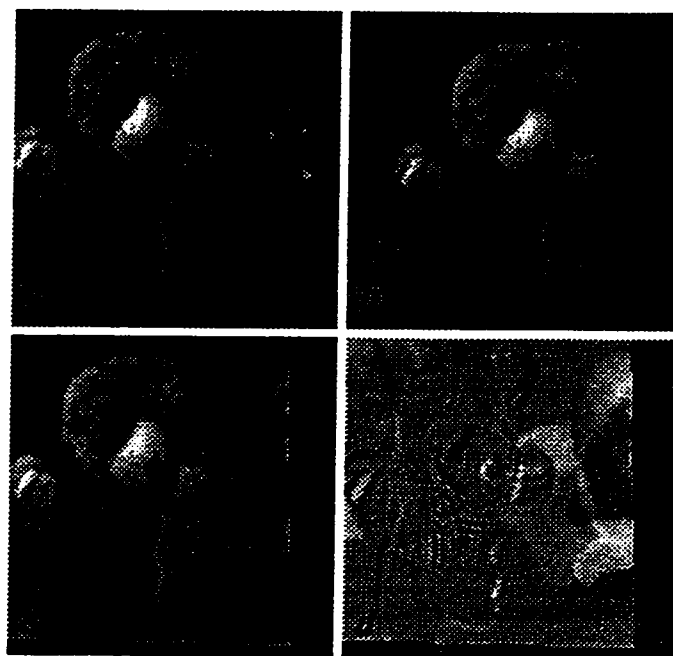
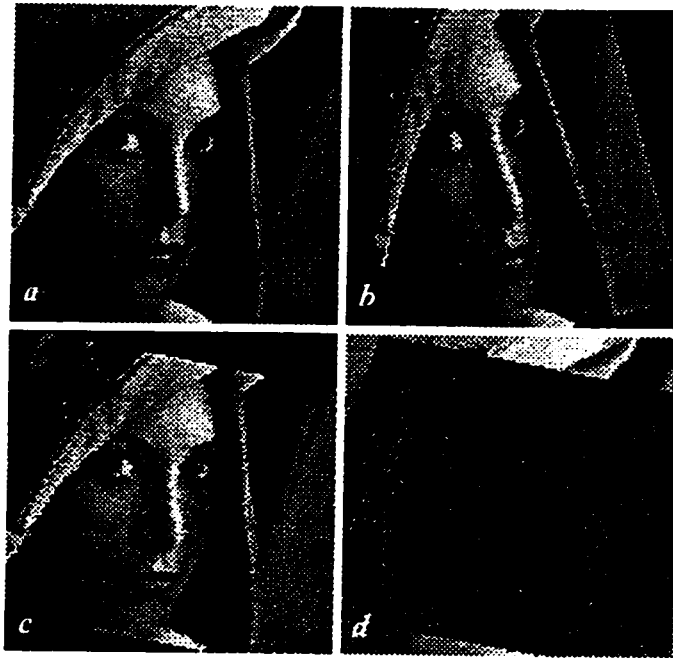


Figure 3.8. *a* and *b*: two different views of a toy scene. *c*: the alignment result of *b* to *a* after 200 time-steps. *d*: difference between *a* and *c*.



*Figure 3.9. a and b: Lenna image and its affine transformation. c: the alignment result of *b* to *a* after 200 time-steps. d: difference between *a* and *c*.*

Chapitre 4

Modèle probabiliste d'appariement

Article: Probabilistic model of multiple dynamic curve matching for semitransparent scene of nonrigid objects

Après l'alignement global, l'appariement entre le modèle déformable et l'image observée permet la segmentation et l'étiquetage de l'image. Le modèle de frontières à multiples contours dynamiques développé dans le présent article reflète la réalité de la superposition des structures d'un objet en projection semitransparente (Mallouche *et al.*, 1997b). Le modèle est appelé *simple* car il comporte uniquement des contours extrémaux. Le cadre utilisé pour développer le critère d'appariement est probabiliste.

4.0 Résumé

Cet article est consacré à l'introduction d'un modèle *simple* à multiples contours dynamiques et au développement du critère probabiliste qui guide la recherche de la solution. L'objectif est de trouver un critère probabiliste à forme distribuée intégrant d'une part une interaction entre les contours et l'image et d'autre part une interaction mutuelle entre les différents contours. Il est aussi souhaitable que le critère puisse assurer un niveau élevé de précision dans la détection et la robustesse au bruit ainsi qu'à la variation du contraste.

On propose la maximisation de la distribution de *pseudo-probabilité conjointe* (MPPC) comme critère de recherche. La distribution de *pseudo-probabilité conjointe* (PPC) peut être vue comme une approximation à la distribution de probabilité *conjointe* globale. La distribution PPC est donnée par le produit des distributions *conjointes* locales.

$$\Phi = \prod_s p(l_{\delta_s}, Y_{\delta_s})$$

où $l = \{l_k, k=1, \dots, K\}$ et Y désignent respectivement la solution et l'image observée. Φ possède des propriétés intéressantes qui justifient notre choix: la forme distribuée permet d'exprimer le critère global en fonction d'interactions locales et par conséquent de pouvoir implanter le processus de recherche sur des calculateurs parallèles; la redondance de la forme locale assure une certaine mesure d'adéquation globale; la dépendance locale de la solution exprimée par l_{δ_s} introduit une mesure d'adéquation locale; la dépendance par rapport au voisinage local Y_{δ_s} permet d'offrir une robustesse élevée par rapport au bruit.

Le critère de MPPC est développé pour un modèle à multiples contours dynamiques en tenant compte principalement de la dualité ligne-région. La fonction du critère est exprimée par le produit d'un terme relié aux données, d'un terme relié au modèle et d'un terme de lissage. Trois formes générales du terme relié aux données sont ensuite discutées. Ces formes traduisent trois différentes partitions du système de voisinage δ_s . La forme retenue possède trois propriétés importantes: la précision et la cohérence de la détection ainsi que la faible répulsion entre les contours voisins sans intersection. Le terme relié au modèle est développé en utilisant une hypothèse d'indépendance locale de la partition des étiquettes de régions $r_{\setminus k, \delta_s} = \{r_{1, \delta_s}, \dots, r_{k-1, \delta_s}, r_{k+1, \delta_s}, \dots, r_{K, \delta_s}\}$ conditionnée par l_{k, δ_s} . Ce terme peut être donné par la

distribution statistique des caractéristiques géométriques locales des segments de contours du modèle. Le terme de lissage est choisi comme la somme pondérée de l'intégration curviligne de la longueur locale et des variations quadratiques de premier et deuxième ordres de l'orientation locale. La fonction d'objectif finale est donnée par le logarithme de la distribution PPC.

L'appariement du modèle probabiliste est testé avec des images synthétiques fortement bruitées. Les objets utilisés dans les expériences sont générés par la superposition semitransparente de formes planaires. Les résultats sont cohérents avec les modèles utilisés, la détection des discontinuités d'intensité est précise, et le processus de recherche est stable et robuste. Des expériences additionnelles confirment l'importance des connaissances relationnelles apportées par le terme relié au modèle dans la stabilisation du processus de recherche.

Les contributions majeures du chapitre sont les suivantes: (i) L'introduction du critère de maximisation de la distribution de *pseudo-probabilité conjointe*. (ii) Le développement du critère avec un modèle *simple* à multiples contours dynamiques représentant la semitransparence du processus de formation de l'image. (iii) L'introduction d'une interaction entre la solution et les connaissances relationnelles du modèle sans compromettre la précision de la détection. (iv) La définition d'une fonction d'objectif finale offrant un contexte robuste, précis et cohérent pour l'appariement du modèle à multiples contours dynamiques avec images de scènes semitransparentes.

**Probabilistic Model of Multiple Dynamic Curve Matching
for Semitransparent Scene of Nonrigid Objects**

Hail Mallouche, Jacques A. de Guise†, Yves Goussard

Biomedical Engineering Institute, Ecole Polytechnique of Montreal
P.O.Box 6079, Station "Centre-Ville," Montreal, Quebec, Canada, H3C 3A7

†Automated Production Engineering, Ecole de Technologie Superieure
1100, Notre-Dame West, Montreal, Quebec, Canada, H3C 1K3

Correspondence should be sent to:

Hail Mallouche

Biomedical Engineering Institute

Ecole Polytechnique of Montreal

P.O.Box 6079, Station "Centre-Ville"

Montreal, Quebec, Canada

H3C 3A7

Phone: (514) 340-4198

Fax: (514) 340-4611

Email: hail@grbb.polymtl.ca

Probabilistic Model of Multiple Dynamic Curve Matching for Semitransparent Scene of Nonrigid Objects

Hail Mallouche, Jacques A. de Guise, Yves Goussard

Abstract. High-level and knowledge-based modeling for scene understanding, organization, and recognition are of great interest in psychophysical and machine vision research. On a lower level, dynamic curves have become an attractive subject in computer vision because of their connectivity and closure properties. Here to fore, proposals were restricted to using one deformable curve, disjoint curves, or elastic models.

In this paper, we present a probabilistic model of multiple dynamic free curve matching with 2-D images. The model is conceived for images from semitransparent scenes such as additive projection imaging – planar radiographical transmission or planar emission imaging. The model is made up of three parts: (i) a data driven term, (ii) a model driven term, and (iii) smoothing constraints. The model represents a hierarchical structure of three processes: lines, regions and relational graph. Thus, an object is modeled as a set of linked subobjects according to a 3-D relational graph that can be projected from a known pose so as to form a 2-D region adjacency graph. The graph represents high-level structural knowledge that reduces the search space dimension by conditioning the *a posteriori* probability to the *a priori* known graph. The resultant objective function is optimized using a descending search method with randomized sampling. Finally, successful results are presented for object matching in semitransparent noisy synthetic scenes.

Index Terms. Multiple dynamic curves, probabilistic models, high-level constrained search, model-based recognition and labeling, semitransparent scene analysis.

4.1 Introduction

Image analysis of semitransparent (ST) scenes is of great interest in medical imaging and in industrial applications (e.g., product inspection). Transparency is due to the penetration capacity of the inspection signal through the material. Semitransparency is generally due to photon attenuation, diffusion or diffraction properties of the investigated material. Usual images are produced under normal illumination conditions using visible light [3]. Therefore, the computer vision community has mostly been interested in usual images from opaque scenes. However, transparent and ST objects exist too. In this paper, we are interested in segmentation and labeling of images from ST scenes.

The occlusion problem is relevant in computer vision [3]. It is due to the opacity of most real objects with respect to visible light. Semitransparency partially reduces this problem. Nevertheless, we are faced with the problem of complexity caused by the superposition of structures. Moreover, images from ST scenes have a low signal to noise ratio (S/N) generally due to photon diffusion.

Consequently, the problem of segmentation and labeling of images from ST scenes are not trivial. It is generally translated as a search problem in the space of possible states. Approaches based on edge detection, feature organization, region merging and labeling, and finally recognition [3] can encounter serious problems when applied to noisy images of ST

scenes. This is basically due to the overfragmented boundaries produced by the edge detector, which dramatically expands the search space. Model-based approaches generally bring additional information in different forms and thus can reduce the search space dimension [19], [7]. These approaches provide good performance when used for controlled scenes as, e.g., in medical and industrial applications.

In this paper, we propose a model-based approach for segmentation and labeling of noisy images from ST scenes. Many kinds of images can be analyzed with our system: medical images like 2-D radiographic transmission projection or 2-D planar emission imaging, industrial images from X-ray inspection and quality control applications, or images of ST objects with respect to visible light. We adopt a stochastic framework that allows us to explore the search space using a single global criterion. The approach integrates a new representation of multiple dynamic curves into a *pseudo-joint* criterion. The objective function (OF) is composed of a data driven term, a model driven term, and a smoothing term. In the model driven term, we use the closure property of the dynamic curves to introduce a high-level interaction with *a priori* relational knowledge of the object.

The paper is organized as follows. The problem statement and related work are presented in Sections 4.2 and 4.3, respectively. In Section 4.4, we introduce the *pseudo-joint* estimator and develop the probabilistic model for matching multiple dynamic curves. In Section 4.5, we develop and discuss the data driven term. Section 4.6 is devoted to the model driven term and the smoothing term. We recapitulate the OF in Section 4.7. The implementation and the optimization method are described in Section 4.8. In Section 4.9, we present and discuss the experimental results of model matching with noisy synthetic images. The approach and the

future research are finally discussed in Section 4.10.

4.2 Problem Statement

Analysis of images from ST scenes received a little of attention either in computer vision or in medical imaging. However, X-ray image analysis can aim at many objectives such as region labeling, object recognition, and 3-D scene reconstruction from a small number of images.

Object recognition from medical X-ray images is not a trivial task even for trained observers. Generally, these images are complex and noisy due to three basic factors. (i) The photon statistics of the X-ray image formation process results in a weak S/N ratio. (ii) The ST superposition of the structures in the 2-D projection and the 3-D geometric complexity of the organs are the main causes of the complexity of X-ray medical images. (iii) The variable and complex background on which the object can be projected further increase the image complexity. Additional difficulties are related to the nature of the imaged objects: (i) The deformable nature of the observed organs precludes using rigidity constraints. Although each human organ owns its specific aspect, it can exhibit a large degree of variability depending on many parameters such as age, gender, race, diseases, pathological situations, life style, etc. (ii) The non-polyhedral geometry of human organs generally yields 2-D projections that generate variant features with respect to the viewing position. Consequently, the early operation of feature extraction usually yields overfragmented boundary elements when applied to noisy and complex images [3] which dramatically increases the search space dimension and renders the visual inverse problem of underconstrained nature. Hence, additional constraints

have to be employed to stabilize the search process.

After feature extraction, feature grouping is used in order to reduce the search space dimension and concentrate the image information [19]. Grouping uses perceptual organization criteria generally expressed as geometric constraints [25], [29], [15], [27]. Nevertheless, in noisy and complex images of superposed structures, low level feature grouping can get trapped in local optima. In order to overcome this difficulty, additional model-based information is needed such as the intrinsic grouping properties of the geometric shapes used in direct matching methods and the relational and structural cues of the model. In fact, relational approaches have been successfully used in computer vision systems for interpretation [26], [32] and shape description [17] of natural images.

Direct matching between the 3-D or 2-D models and the extracted image features has been used for recognizing rigid objects with strong viewpoint invariant features when the initial estimate of the viewing position is close to the real one [7], [19]. However, these methods can encounter serious difficulties when directly applied to noisy images of non-rigid objects with weak viewpoint invariant property as, e.g., X-ray images of human body.

Because a model-based system generally helps reducing the search space dimension, our approach is based on direct matching between the image and the 2-D deformable model of the object [20], [15]. The model is made up of the superposition of multiple dynamic curves that have intrinsic grouping properties. Instead of using rigidity constraints that are inadequate for deformable object, or elastic constraints that bias the precision of detection toward the unstressed state of the model, we introduce the structural knowledge of the model as a high-level interaction in the matching process.

4.3 Related Work

We review three research subjects related to our work: dynamic curves, elastic models, and template matching.

A dynamic curve (DC) is a deformable continuous closed curve. The use of DC in computer vision can inherently overcome two important difficulties in feature grouping: edge organization and region merging [25], [29], [19]. This is due to their connectivity and closure intrinsic properties [16], [31]. The connectivity implicitly incorporates the continuity and proximity criteria that are extensively used in perceptual organization [25], [29], [23]. Curve closure reflects the line-region duality and includes implicit region merging property. These intrinsic properties of DC dramatically reduce the search space dimension and make DC of great interest in computer vision [31].

Dynamic curves (DCs) were mostly used in computer vision systems for controlled scenes such as, e.g., medical scenes [13], [28], [30], [31]. Proposals were mostly restricted to segmenting images of opaque scenes using one DC [33], [34], [39]. DCs have also been used in signal matching [39], contour extraction of moving objects [11], real time tracking [9][13][16] and stereo matching [16]. DC has mostly been modeled as a snake [16] that seeks an equilibrium state by optimizing the sum of internal and external energy functions [33]. The internal energy generally represents smoothing constraints. The external energy is a data driven force that deforms the curve so as to match it with the image. Snakes have been used for matching with images of the real world [11], [9], binary images [17], [22], medical images [13], [28], [31], and handwritten images [18]. In order to achieve single criterion matching

[28], [30], deterministic methods, e.g., variational [34], [39] and descending ones [28], [30], and stochastic methods [31], [14] have been used. Generally, these methods have been limited to one DC or disjoint DCs applied to simple images, and required that the initial state be close to the real solution to ensure safe convergence.

For complex images, model-based methods such as, e.g., elastic matching are almost unavoidable [2], [35], [36]. These methods generally solve the dynamic equation of the model subjected to external forces. The elastic energy measures the deviation of the metric tensor of the model inner filling material from its initial state. The model recovers its initial state when external forces vanish. The solution results from the balance between external and internal forces. The elastic component introduces a permanent bias toward the unstressed state. Therefore, the elasticity level controls the tradeoff between data faithfulness and robustness. The precision of detection can be heavily compromised with noisy and low contrast images or when the initial state is relatively far from the real solution [2]. In fact, the precision can be improved by substantially decreasing the elasticity of the model which also reduces the model to a simple initial state. Thus far, such elastic models have not integrated any structural knowledge. Similarly, parametric models have been used [30], [18]. The precision is also altered by the tradeoff between the external forces and *a priori* parameters distribution.

Intensity-based methods, e.g., template matching, have been investigated for image restoration and atlas matching [1], [14]. These methods encounter serious difficulties with intensity changing objects and with varying background. Even with powerful optimization methods such as stochastic differential equations (SDE) the convergence is not guaranteed

when the initial state is relatively far from the real solution. In addition, the solution is compromised by the SDE boundary conditions.

Finally, structural information seems important for recognition and interpretation in human vision [40], [23]. This may explain the success of the approaches based upon relational models for image interpretation [26], [3], [12]. In addition, relational and structural knowledge of the considered scene can be helpful for low level feature organization. Consequently, we believe that such high-level information can provide structural constraints in a multiple curve matching approach.

In this paper, we wish to develop a system for precise detection of intensity discontinuities and region labeling in noisy and complex images from ST scenes with low and variable contrast. Model-based systems are almost unavoidable for labeling and recognition tasks. However, elastic and intensity-based models are inconvenient for the type of images we want to analyse. We here propose a model-based probabilistic approach for matching with multiple free²⁸ dynamic curves. We assume that viewing position (projection pose) is approximately known and that the model integrates high-level relational knowledge. Inner-outer region superposition of the closed boundary curves are related by a 2-D relational graph which can be seen as the projection of the 3-D representation of the object. The detection and labeling are performed by matching the 2-D object model with the observed image. The matching is guided by the optimization of a global probabilistic criterion. Additionally, we integrate a high-level relational knowledge in the search process without altering the precision of the detection.

²⁸

Models free of material (can be freely deformed in all directions) opposed to elastic models.

4.4 Matching Multiple Dynamic Curves

Matching between the model and the data can be described in a stochastic framework. Many probabilistic estimators were proposed for image analysis and scene understanding. However, three basic characteristics are generally desired for the probabilistic estimator: (i) coherence of the estimate (generated solution); (ii) precision and (iii) robustness of detection; (iv) distributed form of the global criterion for simplicity and computational efficiency.

The *coherence* means that the general shape of the solution is consistent with the *a priori* known model and the observed data and that the solution detects all the constituting parts of the object in the image. In Section 4.9, we show examples of coherent solutions and incoherent solutions. The *precision* means that the detected boundaries are positioned on the real location of the intensity discontinuities of the image. These discontinuities generally represent the projection of the silhouette and the surface discontinuities of the imaged object. The *robustness* means that the estimate degrades gracefully with the increase of noise level in the image and modeling error of the initial state.

The *maximum a posteriori* (MAP) criterion has been widely used because it expresses the general scene understanding and the reconstruction problem in a single objective function [10], [24]. Let Y be the observed image which is assumed to be a realization from a random field \mathbf{Y} . \mathbf{Y} is defined in terms of an underlying random field \mathbf{I} (scene). Let I be a realization of \mathbf{I} . MAP estimation consists of maximizing the posterior probability $p(I=I | \mathbf{Y}=Y)$ with respect to I . In order to lighten the notations, $p(I | Y)$ will denote $p(I=I | \mathbf{Y}=Y)$ whenever unambiguous. Often, $p(I | Y)$ is of complex nature and of high dimensionality. This makes

it difficult to derive its explicit form and to visualize its behavior [4]. In addition, MAP global solutions can lead to local inconsistencies. Instead, the *marginal posterior modes* (MPM) and the *marginal and conditional posterior* (MCP) have been proposed [4], [24] to correct for the MAP local inconsistency. However, both MPM and MCP lack global consistency and robustness.

We now propose the *maximum pseudo-joint* (MPJ) criterion for the multiple DCs matching problem. We show that MPJ criterion offers interesting properties relatively to other estimators.

4.4.1 Maximum Pseudo-Joint Criterion

Let L be the basic discrete lattice. The random field Y corresponding to the observed image Y is defined on the lattice L^Y that is similar to L . Let I be a realization of the underlying solution random field I defined on the lattice L^I that is also similar to L . Let the operator $|\cdot|$ denotes the set or the subset cardinal. We note $c = |L|$. Let s be a discrete site of L . Let δs denotes a neighborhood system centered on s , $s \in \delta s \subset L$.

Many researchers expressed $p(I|Y)$ in simple forms by introducing different assumptions [4], [6], [5], [10]. These forms were mostly distributed and formulated as products of local probabilities. Conditional independence assumptions have been used for modeling textured images [10], as:

$$p(I|Y) = \prod_{s \in L} p(I_s | Y_{\delta s}) \quad (4.1)$$

where $p(I_s | Y_{\delta s})$ denotes the conditional probability $p(I_s = I_s | Y_{\delta s} = Y_{\delta s})$. This assumption

is basically not valid, but is practically useful when the spatial dependency of \mathbf{I} can be included in the model component $p(\mathbf{l})$. The *pseudo-likelihood* was proposed as a practical and convenient procedure to approximate the *likelihood* distribution [6], [5], [4]. Here, \mathbf{I} is assumed locally and conditionally independent, $p(\mathbf{l}) = \prod_{s \in L} p(l_s | l_{\delta s \setminus s})$ where

$\delta s \setminus s = \{t \in \delta s, t \neq s\}$. In the *pseudo-likelihood* distribution, the local conditioning and the product on L ensure some level of local and global consistency of the solution with the model, respectively. In fact, we have been inspired by the *pseudo-likelihood* criterion [6], [5] and the assumption (4.1) [10] to propose the *maximum pseudo-joint* criterion [41] for our multiple curve matching problem. The MPJ estimate is defined by:

$$\mathbf{l}^* = \arg \max_{\mathbf{l}} \left\{ \prod_{s \in L} p(l_{\delta s}, Y_{\delta s}) \right\} \quad (4.2)$$

Local joint probability $p(l_{\delta s}, Y_{\delta s})$ can be equivalently replaced in (4.2) by the local conditional probability $p(l_{\delta s} | Y_{\delta s})$. In some manner, the MPJ criterion integrates both local conditionally independent assumptions with respect to \mathbf{I} and \mathbf{Y} . The functional of (4.2) is the *pseudo-joint* distribution. It can be seen as the product of the local joint distributions [41]. In addition, the integration of the local dependency of $\mathbf{l}_{\delta s}$ and $\mathbf{Y}_{\delta s}$ in the same distribution creates a larger local interaction between the solution state and the observed image comparatively to the local term $p(l_s | Y_{\delta s})$ of (4.1). $p(l_{\delta s} | Y_{\delta s})$ can also be seen as a neighborhood-to-neighborhood matching between the solution state and the observed image. This will allow us to introduce,

in the final criterion, a local measurement of mutual coherence between the multiple curves of the desired solution and a consistency measurement between the solution and the observed image. However, this will yield a higher complexity of the model and the optimization method. We note that the *maximum joint* (MJ) solution given by maximizing the *joint* distribution $p(l, Y)$ and the MAP solution are the same. The MPJ solution can also be seen as an approximation to the MJ and similarly to the MAP solution.

In order to see the advantages of MPJ criterion, we recall $p(l_{\delta s} | Y_{\delta s}) \propto p(Y_{\delta s} | l_{\delta s}) \cdot p(l_{\delta s})$ that is equal to the *local-likelihood* distribution times *a priori local* distribution of the solution. When δs is reduced to s , we write $p(l_{\delta s} | Y_{\delta s}) \propto p(Y_s | l_s) \cdot p(l_s) \cdot p(Y_{\delta s} | l_{\delta s})$ can be managed to provide a more reliable estimation of the image-to-solution adequacy measurement because it uses a larger quantity of information in a spatially organized neighborhood δs comparatively to $p(Y_s | l_s)$. In addition and comparing with $p(l_s)$, $p(l_{\delta s})$ can bring a larger *a priori* spatial knowledge about the solution into the search process. These advantages will allow us to develop an explicit expression of MPJ integrating interesting properties.

4.4.2 Model for Multiple Dynamic Curves

In the following, we develop a probabilistic model for matching multiple DC with images of a ST scene. We first introduce some notations and assumptions. We assume that our prior model for ST scenes is composed of K continuous closed DCs denoted $C = \{C_1, \dots, C_K\}$. Let

$l = \{l_1, \dots, l_K\}$ be a realization of the multiple line random field $\mathbf{l} = \{l_1, \dots, l_K\}$. We also assume that each closed curve cannot be self-crossed. Let $\mathbf{l}_{k,s}$ be the line random field \mathbf{l}_k at the site s . Each $\mathbf{l}_k = \{l_{k,1}, \dots, l_{k,c}\}$, $k=1, \dots, K$, is the set of c binary sites defined on the discrete lattice L_k^l , $k=1, \dots, K$, similar to L . l_k is a realization of \mathbf{l}_k and is referred to as the k th line. Let $l_{\setminus k} = l \setminus l_k = \{l_1, \dots, l_{k-1}, l_{k+1}, \dots, l_K\}$, $l_{\delta s} = \{l_t \in l \mid t \in \delta s\}$, $l_{k,\delta s} = \{l_{k,t} \in l_k \mid t \in \delta s\}$, and $l_{k,\delta s s'} = \{l_t \in l_{k,\delta s}, t \neq s\}$. The continuity, the closure and the self uncrossing properties of each line l_k , $k=1, \dots, K$, allow us to define a dual region random field $\mathbf{r}_k = \{r_{k,1}, \dots, r_{k,c}\}$. Let $r_k = \{r_{k,1}, \dots, r_{k,c}\}$ denotes a realization of \mathbf{r}_k . (l_k, r_k) is referred to as the line-region duality meaning that one realization can be inferred from the other by direct operations. As \mathbf{l}_k , region \mathbf{r}_k , $k=1, \dots, K$ is binary and defined on discrete lattice L_k^r , $k=1, \dots, K$, similar to L . The realization $r = \{r_1, \dots, r_K\}$ is referred to as the set of regions. Region r_k describes the inner and outer topological regions of the line l_k corresponding to the closed curve C_k . Fig. 4.1 shows the structure of three closed curves. The passage from line l_k to region r_k is straightforward if l_k implicitly contains the curve orientation information – clockwise orientation if the inner region is the right-hand region of the local line segment. Let $r_{\setminus k} = r \setminus r_k = \{r_1, \dots, r_{k-1}, r_{k+1}, \dots, r_K\}$ and $r_{\setminus k, \delta s} = \{r_{\setminus k, t} \in r_{\setminus k} \mid t \in \delta s\}$. The unknown we seek to determine is the set l of K lines. The global criterion is given by the MPJ estimator where L

is replaced by the set of binary latices $L = \{L_1^l, \dots, L_K^l\}$.

Let the functional of (4.2) be denoted by Φ . The redundancy defined on the set $\{L_1^l, \dots, L_K^l\}$ in Φ is essential for the curve mutual interactions in both the data driven and the model driven terms as will be shown in Sections 4.5 and 4.6, respectively. $p(l_{\delta s}, Y_{\delta s})$ reveals the local Markovian property of the model and the local dependency of the solution onto the observed image. Now, we can split the product in Φ into two operations defined on $k=1, \dots, K$, and $s \in L_k^l$. Then, we rearrange the local $\{l_{\delta s}\}$ for each $k=1, \dots, K$ as $l_{\delta s} = \{l_{1,\delta s}, \dots, l_{K,\delta s}\} = \{l_{k,\delta s}, l_{\setminus k,\delta s}\} = \{l_{k,\delta s}, r_{\setminus k,\delta s}\}$ where we substitute $r_{\setminus k,\delta s}$ for $l_{\setminus k,\delta s}$ using the line-region duality of the closed DC which translates the equivalency between of the information contained in $l_{\setminus k,\delta s}$ and in $r_{\setminus k,\delta s}$. Introducing this in Φ yields:

$$\Phi = \prod_{k=1, \dots, K} \prod_{s \in L_k^l} p(l_{k,\delta s}, r_{\setminus k,\delta s}, Y_{\delta s}) \quad (4.3)$$

Let $p_{k,s}$ denotes the inner term of (4.3). The arrangement $l_{\delta s} = \{l_{k,\delta s}, r_{\setminus k,\delta s}\}$ is useful to express $p_{k,s}$ as a product of three conditional probabilities representing three different interactions. It also allows each line l_k to interact with the region $r_{\setminus k}$ of the remaining lines $l_{\setminus k}$. In fact, $r_{\setminus k}$ will be used to partition the local neighborhood δs which will allows us to incorporate some benefic inter-curve local properties as we show in Section 4.5. The term $p_{k,s}$ can be decomposed as follows:

$$\begin{aligned}
p_{k,s} &= p(l_{k,\delta_s}, r_{\setminus k,\delta_s}, Y_{\delta_s}) \\
&= p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s}) \cdot p(l_{k,\delta_s}, r_{\setminus k,\delta_s}) \\
&= p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s}) \cdot p(r_{\setminus k,\delta_s} | l_{k,\delta_s}) \cdot p(l_{k,\delta_s})
\end{aligned} \tag{4.4}$$

Moreover, we replace l_{k,δ_s} by $\{l_{k,s}, l_{k,\delta_s \setminus s}\}$ in $p(l_{k,\delta_s})$ and we decompose it as

$p(l_{k,\delta_s}) = p(l_{k,\delta_s \setminus s} | l_{k,s}) p(l_{k,s})$. Here, we assume that binary variable $l_{k,s}$ is uniformly

distributed. $p_{k,s}$ can then be rewritten as:

$$p_{k,s} \propto p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s}) \cdot p(r_{\setminus k,\delta_s} | l_{k,\delta_s}) \cdot p(l_{k,\delta_s \setminus s} | l_{k,s}) \tag{4.5}$$

The MPJ solution is produced by the criterion (4.2) combined to expressions (4.3) and (4.5). The terms in the RHS of (4.5) combined with the product in (4.2) on the set of lattices $\{L_1^l, \dots, L_K^l\}$ have the following interpretations: (i) The first term measures the consistency of the solution with the observed image. It is referred to either as the image formation part or as the data driven term, or the *likelihood* distribution of the image Y given l . (ii) The second term measures the self-consistency of the solution and can incorporate high-level interactions with the model-based knowledge. It is referred to as the model driven term. (iii) The third term can contain *a priori* smoothing constraints. We now give more precise expression for each term.

4.5 Data Driven Term

We now propose a general form for $p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s})$ of (4.5). $p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s})$ measures the consistency between the local line l_{k,δ_s} and Y_{δ_s} assuming that local regions $r_{\setminus k,\delta_s}$

corresponding to the remaining lines are already known. This general form is compared with two simpler forms in order to show that it can incorporate desired properties: robustness with respect to noise, precision of detection²⁹, coherence of the solution with the image, and weak repulsive behavior between non crossing curves of the solution. The coherence is essential for free form deformable model in order to prevent that the different curves behave inconsistently regarding the other curves. The repulsive property is important at the beginning of the search because it helps preventing that many curves converge together to the same discontinuities. The basic idea is to manipulate the local neighborhood system δs by partitioning it into different subsystems defined on the local labeling region $r_{\setminus k}$ in order to create the above-mentioned properties.

4.5.1 General Form

Let $in(l_k)=in(r_k)=in(k)$ and $out(l_k)=out(r_k)=out(k)$ denote the inner and outer regions of the closed curve l_k , respectively. $l_{k,\delta s}$ splits the local neighborhood system δs centered on site s into two parts, $\delta s_{in(k)}$ and $\delta s_{out(k)}$, where $\delta s_{in(k)}=\{t \in \delta s \mid r_{k,t} \in in(k)\}$ and $\delta s_{out(k)}=\{t \in \delta s \mid r_{k,t} \in out(k)\}$. In the same manner, $r_{\setminus k,\delta s}$ can be partitioned on $\{\delta s_{in(k)}, \delta s_{out(k)}\}$ into subsets $r_{\setminus k,\delta s_{in(k)}}=\{r_{\setminus k,t} \mid t \in \delta s_{in(k)}\}$ and $r_{\setminus k,\delta s_{out(k)}}=\{r_{\setminus k,t} \mid t \in \delta s_{out(k)}\}$. Thus, we can write $r_{\setminus k,\delta s_{IO(k)}}=\{r_{\setminus k,\delta s_{in(k)}}, r_{\setminus k,\delta s_{out(k)}}\}$.

²⁹

Intensity discontinuities.

Let $a = \{a_1, \dots, a_m\}$ be the set of local labels defined on $r_{\setminus k, \delta s}$. Each $a_j, j = 1, \dots, m$, is a $(K-1)$ -ary label that translates the superimposition of the $(K-1)$ binary regions constituting $r_{\setminus k}$. We regroup the sites having similar labels on $r_{\setminus k, \delta s}$. This yields the following partition:

$$r_{\setminus k, \delta s} = \left\{ \left\{ r_{\setminus k, \delta s}^{(a_1)} \right\}, \dots, \left\{ r_{\setminus k, \delta s}^{(a_m)} \right\} \right\} \quad (4.6)$$

where each site $s \in r_{\setminus k, \delta s}^{(a_j)}$ has the label a_j in the region $r_{\setminus k}$. Equally, $\delta s^{(a_j)}$ represents the set

of sites of δs labeled with a_j in $r_{\setminus k, \delta s}$. Thus, we have $r_{\setminus k, \delta s}^{(a_j)} = \{r_{\setminus k, t} \mid t \in \delta s; r_{\setminus k, t} = a_j\}$ and

$\delta s^{(a_j)} = \{t \in \delta s \mid r_{\setminus k, t} = a_j\}$. Then, we write $p(Y_{\delta s} \mid l_{k, \delta s}, r_{\setminus k, \delta s}) = p\left(Y_{\delta s} \mid l_{k, \delta s}, \left\{r_{\setminus k, \delta s}^{(a_1)}, \dots, r_{\setminus k, \delta s}^{(a_m)}\right\}\right)$.

The idea here is to express the *local likelihood* distribution $p(Y_{\delta s} \mid l_{k, \delta s}, r_{\setminus k, \delta s})$ as a product of the set of *sub-local likelihood* distributions $p\left(Y_{\delta s}^{(a_j)} \mid l_{k, \delta s}, r_{\setminus k, \delta s}^{(a_j)}\right), j = 1, \dots, m$.

$$p(Y_{\delta s} \mid l_{k, \delta s}, r_{\setminus k, \delta s}) \propto \prod_{j=1, \dots, m} p\left(Y_{\delta s}^{(a_j)} \mid l_{k, \delta s}, r_{\setminus k, \delta s}^{(a_j)}\right) \quad (4.7)$$

where each *sub-local* distribution is estimated from the corresponding *sub-local* neighborhood. Recalling the partition $\{\delta s_{in(k)}, \delta s_{out(k)}\}$ generated by the local line segment $l_{k, \delta s}$

over δs , as shown in Fig. 4.2, we can directly see that it also has the effect of partitioning

each $r_{\setminus k, \delta s}^{(a_j)}$ into $r_{\setminus k, \delta s_{IO(k)}}^{(a_j)} = \left\{r_{\setminus k, \delta s_{in(k)}}^{(a_j)}, r_{\setminus k, \delta s_{out(k)}}^{(a_j)}\right\}, j = 1, \dots, m$. Then, introducing the partition

$r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}$ in (4.7) yields:

$$\begin{aligned}
p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s}) &\propto \prod_{j=1,\dots,m} p\left(Y_{\delta_s^{(a_j)}} | \left\{ r_{\setminus k,\delta_{s_{in}(k)}}^{(a_j)}, r_{\setminus k,\delta_{s_{out}(k)}}^{(a_j)} \right\}\right) \\
&\propto \prod_{j=1,\dots,m} p\left(Y_{\delta_{s_{in}(k)}^{(a_j)}}, Y_{\delta_{s_{out}(k)}^{(a_j)}} | \left\{ r_{\setminus k,\delta_{s_{in}(k)}}^{(a_j)}, r_{\setminus k,\delta_{s_{out}(k)}}^{(a_j)} \right\}\right)
\end{aligned} \tag{4.8}$$

where $Y_{\delta_{s_{in}(k)}^{(a_j)}}$ and $Y_{\delta_{s_{out}(k)}^{(a_j)}}$ denote two subsets of Y_{δ_s} defined as

$Y_{\delta_{s_{in}(k)}^{(a_j)}} = \{Y_t | t \in \delta_s^{(a_j)}; t \in in(k)\}$ and $Y_{\delta_{s_{out}(k)}^{(a_j)}} = \{Y_t | t \in \delta_s^{(a_j)}; t \in out(k)\}$, respectively. Let

$\bar{Y}_{\delta_{s_{in}(k)}^{(a_j)}}$ and $\bar{Y}_{\delta_{s_{out}(k)}^{(a_j)}}$ be the local image mean values computed on the sublattices $\delta_{s_{in}(k)}^{(a_j)}$ and

$\delta_{s_{out}(k)}^{(a_j)}$, respectively. We now propose the following two-step approximation:

$$\begin{aligned}
p(Y_{\delta_s} | l_{k,\delta_s}, r_{\setminus k,\delta_s}) &\approx \prod_{j=1,\dots,m} p\left(Y_{\delta_{s_{in}(k)}^{(a_j)}}; Y_{\delta_{s_{out}(k)}^{(a_j)}}\right) \\
&\approx \prod_{j=1,\dots,m} p\left(\bar{Y}_{\delta_{s_{in}(k)}^{(a_j)}}; \bar{Y}_{\delta_{s_{out}(k)}^{(a_j)}}\right)^{\rho_{s,j}}
\end{aligned} \tag{4.9}$$

(4.9) can be seen as a *sub-local* patching of the image formation process using some locally uniform template $r_{\setminus k,\delta_s}^{(a_j)}$. The explicit form of (4.9) is discussed in Section 4.8. The first step

in (4.9) can be justified by the worthlessness of defining a model template and the difficulty of precisely determining the image formation process for background and contrast varying images as, e.g., medical X-ray images. Hence, in the data driven term, we only maintain the local topological information of the partition $\{\delta_{s_{in}(k)}^{(a_j)}, \delta_{s_{out}(k)}^{(a_j)}\}$ defined by the superposition

of the regions. The second step is equivalent to two approximations. (i) If we can assume that

the image formation process can be modeled using a template and an i.i.d. additive noise as in (4.15), then the probability term $p\left(Y_{\delta s_{in(k)}^{(a_j)}}; Y_{\delta s_{out(k)}^{(a_j)}}\right)$ can be written as a product of probabilities defined over the sites of $\delta s^{(a_j)}$. So, we see that $\rho_{s,j}$ can be chosen for example equal to $|\delta s^{(a_j)}|$. In fact, $\rho_{s,j}$ represents the statistical weight of (or the confidence in) the estimate of the *sub-local* distribution $p\left(Y_{\delta s^{(a_j)}} \mid r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right)$. $\rho_{s,j}$ is a function of the number of samples used to estimate the distribution. (ii) Instead of using the real value $Y_{t \in \delta s_{in(k)}^{(a_j)}}$ ($Y_{t \in \delta s_{out(k)}^{(a_j)}}$) at $t \in \delta s_{in(k)}^{(a_j)}$ ($t \in \delta s_{out(k)}^{(a_j)}$) in $p\left(Y_{\delta s_{in(k)}^{(a_j)}}; Y_{\delta s_{out(k)}^{(a_j)}}\right)$ we replace each $Y_{t \in \delta s_{in(k)}^{(a_j)}}$ ($Y_{t \in \delta s_{out(k)}^{(a_j)}}$) by the sub-local mean value $\bar{Y}_{\delta s_{in(k)}^{(a_j)}}$ ($\bar{Y}_{\delta s_{out(k)}^{(a_j)}}$). This allows us to reduce the sensitivity of the estimate of the local probability with respect to noise and therefore to give to it a robust behavior. Note that each $\bar{Y}_{\delta s_{in(k)}^{(a_j)}}$ and $\bar{Y}_{\delta s_{out(k)}^{(a_j)}}$ is a single site value while $Y_{\delta s_{in(k)}^{(a_j)}}$ and $Y_{\delta s_{out(k)}^{(a_j)}}$ are defined on $|\delta s_{in(k)}^{(a_j)}|$ and $|\delta s_{out(k)}^{(a_j)}|$ sites, respectively. The local mean can also be replaced by any weighted local mean.

Simple forms of the data driven term can also be derived. Let α_s be the label of the site s in $r_{\setminus k}$. The second form is obtained by assuming that the *local likelihood* $p\left(Y_{\delta s} \mid l_{k, \delta s}, r_{\setminus k, \delta s}\right)$

can be estimated by the *sub-local likelihood* defined on the *sub-local* part $r_{\setminus k, \delta_s}^{(a_s)}$ of $r_{\setminus k, \delta_s}$. This

yields:

$$p(Y_{\delta_s} | l_{k, \delta_s}, r_{\setminus k, \delta_s}) \propto p\left(\bar{Y}_{\delta_{s_{in(k)}}}^{(a_s)}; \bar{Y}_{\delta_{s_{out(k)}}}^{(a_s)}\right)^{\rho_s} \quad (4.10)$$

The third form is the simplest. Here, we assume that $p(Y_{\delta_s} | l_{k, \delta_s}, r_{\setminus k, \delta_s})$ depends only on the line l_k . Then, using the partition $\{\delta_{s_{in(k)}}, \delta_{s_{out(k)}}\}$ defined by l_{k, δ_s} , we propose the third form as:

$$p(Y_{\delta_s} | r_{\setminus k, \delta_s}, l_{k, \delta_s}) \propto p\left(\bar{Y}_{\delta_{s_{in(k)}}}; \bar{Y}_{\delta_{s_{out(k)}}}\right)^{\rho} \quad (4.11)$$

where $Y_{\delta_{s_{in(k)}}} = \{Y_t | t \in \delta_{s_{in(k)}}\}$ and $Y_{\delta_{s_{out(k)}}} = \{Y_t | t \in \delta_{s_{out(k)}}\}$ are two disjoint subsets of Y_{δ_s} .

4.5.2 Local Data Driven Term - Discussion

We now discuss the behavior of the general forms of the data driven term. Figures 4.3 and 4.4 illustrate schemes of crossing curves and non-crossing curves, respectively. We assume that curve l_h perfectly matches the observed image Y and thus has reached its optimal position. Conversely, we assume that l_k will be progressively deformed or positioned in order to match Y . In addition, we assume that any $p(Y_{s \in in(k)}; Y_{t \in out(k)})$ is an increasing function of

$Y_{s \in in(k)} - Y_{t \in out(k)}$. Note that $|r_{\setminus k, \delta_s}^{(a_j)}| = |\delta_s^{(a_j)}| = |\delta_{s_{in(k)}}^{(a_j)}| + |\delta_{s_{out(k)}}^{(a_j)}|$. We now verify which

of expressions (4.9), (4.10) and (4.11) offers the robustness, precision of detection, coherence

and weak repulsive properties without undesired behaviors such as implicit orthogonality or smoothing constraints.

We recall (4.9) in order to examine some possible expressions for $\rho_{s,j}$. Each expression exhibits some intrinsic properties that directly modify the nature of the state space. Let $\rho_{s,j}$ equal $\iota_{1,j} = 2 \min \left\{ |\delta s_{in(k)}^{(a_j)}|, |\delta s_{out(k)}^{(a_j)}| \right\}$ or $\iota_{2,j} = |\delta s_{in(k)}^{(a_j)}| + |\delta s_{out(k)}^{(a_j)}|$ where $\iota_{2,j}$ is set to zero when either the inner part or the outer part of $\delta s^{(a_j)}$ is empty. $\iota_{1,j}$ introduces intrinsic smoothing effect, strong mutual repulsive influence on any two non-crossing neighboring curves, and orthogonality constraint on the crossing curves. Fig. 4.4 shows a typical situation, where l_k tends to take an orthogonal position to the direction of l_h in order to maximize $\iota_{1,j}$.

On the other hand, $\iota_{2,j}$ does not introduce any smoothing effect. In fact, $\iota_{2,j}$ is equal to $|\delta s|$ when s belongs to any isolated segment. In addition, $\iota_{2,j}$ exerts a weak repulsive effect on the non-crossing neighboring curves. Fig. 4.4 shows a typical situation where we can assume that l_k and l_h have rigid shapes and l_k is moved closer to l_h starting from a position where $|\delta s_{out(k)}^{(a_j)}| \leq |\delta s_{in(k)}^{(a_j)}|$. $\iota_{1,j}$ will decrease as $2 |\delta s_{out(k)}^{(a_j)}|$ while $\iota_{2,j}$ will vary as $|\delta s^{(a_j)}|$.

Obviously, if $\rho_{s,j} = \frac{\iota_{1,j}}{|\delta s|}$ then $\rho_{s,j}$ has the same properties of $\iota_{1,j}$. However, if

$\rho_{s,j} = \iota_{s,j} / \sum_{i=1,\dots,m} \iota_{s,i}$, the smoothing, repulsive, and orthogonality properties are neutralized,

and hence ρ_j will have almost the same behavior using either $\iota_{1,j}$ or $\iota_{2,j}$.

We now recall the second form (4.10) where $\rho_{s,j}$ can take the same expressions as $\rho_{s,j}$ in (4.9). (4.10) has a major drawback for crossing-curve schemes where $|r_{\setminus k, \delta s_{in(k)}}^{(a_j)}|$ and $|r_{\setminus k, \delta s_{out(k)}}^{(a_j)}|$ decrease monotonically as the site $s \in l_k$ goes closer to the crossing point s_c (Fig.

4.3). This yields a progressive increase in the sensitivity of $\bar{Y}_{\delta s_{in(k)}}^{(a_j)}$ and $\bar{Y}_{\delta s_{out(k)}}^{(a_j)}$ with respect to noise. Oppositely, the third form (4.11) does not suffer from this drawback but instead it has incoherent behavior because it does not consider the remaining $r_{\setminus k, \delta s}$. If we assume for Figures 4.3 and 4.4 that l_h matches with a high contrast discontinuity in Y , while l_k tries to match with a low contrast discontinuity close to l_h . For non-crossing curves (Fig. 4.4), the curve l_h exerts a strong repulsive effect on l_k which compromises the precision. On the other hand, l_k can converge to the same intensity boundary of l_h if it crosses l_h (Fig. 4.3) yielding an inconsistent result. Contrarily, (4.9) yields a coherent and precise matching for both crossing and non-crossing situations. Consequently, (4.9) exhibits the most interesting properties and particularly when combined with $\rho_{s,j} = \frac{|r_{\setminus k, \delta s}^{(a_j)}|}{|\delta s|}$.

We empirically show how the MJP criterion allows us to develop an explicit expression of the funtionnal by incorporating interesting local behaviour.

4.6 Model Driven Term

In this section, we consider both the model driven term $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$ and the smoothing term $p(l_{k, \delta_s} | l_{k, s})$. We apply the label partition $\{r_{\setminus k, \delta_s}^{(a_1)}, \dots, r_{\setminus k, \delta_s}^{(a_m)}\}$ defined on $r_{\setminus k, \delta_s}$, and previously used for the data driven term $p(Y_{\delta_s} | r_{\setminus k, \delta_s}, l_{k, \delta_s})$, to the probability term $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$ of (4.5). We now introduce an independence assumption given by $p(\{r_{\setminus k, \delta_s}^{(a_1)}, \dots, r_{\setminus k, \delta_s}^{(a_m)}\} | l_{k, \delta_s}) = \prod_{j=1, \dots, m} p(r_{\setminus k, \delta_s}^{(a_j)} | l_{k, \delta_s})$. This translates the mutual independence of the partition (4.6) conditionally to l_{k, δ_s} . Though the dependency between the regions $r_{\delta_s}^{(a_1)}, \dots, r_{\delta_s}^{(a_m)}$ exists through the relational graph $G_{0, \theta}$, the assumption seems reasonable since the dependency between $\{l_k, k=1, \dots, K\}$, and consequently between $\{r_k, k=1, \dots, K\}$, is preserved and expressed in the self-consistency term $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$ of (4.5). In this instance, this mutual independency assumption, defined on the partition $\{r_{\setminus k, \delta_s}^{(a_1)}, \dots, r_{\setminus k, \delta_s}^{(a_m)}\}$ and conditioned by l_{k, δ_s} , is used to express the model-solution consistency measurement. We recall that each site $s \in r_{\setminus k, \delta_s}^{(a_j)}$ is labeled by a_j in $r_{\setminus k}$. Now, let $p(a_j | l_{k, \delta_s}) = p(s \in r_{\setminus k, \delta_s}^{(a_j)} | l_{k, \delta_s})$.

Then, we rewrite $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$ as follows:

$$p(r_{\setminus k, \delta_s} | l_{k, \delta_s}) = \prod_{j=1, \dots, m} p(a_j | l_{k, \delta_s})^{\rho_{s,j}} \quad (4.12)$$

$\rho_{s,j}$ is the same as for the data driven term. However, the choice of the expression of $\rho_{s,j}$ is different from the data-solution case. In fact, the choice of $\rho_{s,j}$ does not significantly affect the behavior of (4.12). $p(a_j | l_{k, \delta_s})$ can be approximated by $p(a_j, v_{k,s})$, where $v_{k,s}$ is a set of local attributes of the local line segment l_{k, δ_s} . These attributes should remain of local nature to ensure the local interaction decomposition of the global criterion. However, $p(a_j | l_{k, \delta_s})$ is a model driven source of information about curve l_k crossing the region labeled by a_j in regions $r_{\setminus k}$. Consequently, l_{k, δ_s} can be directly replaced in $p(a_j | l_{k, \delta_s})$ by its local descriptors as, e.g., $v_{k,s} = \{\theta_{k,s}, d\theta_{k,s}, d^2\theta_{k,s}, \dots, d^P\theta_{k,s}\}$ where $d^n\theta$ is the n th curvilinear derivative of the angular local direction. Hence, $p(a_j, v_{k,s})$ is a P parameters sample of the representation $\{p_k^{(a_j)}(d^n\theta), n=1, \dots, P\}$, where $p_k^{(a_j)}(d^n\theta)$ can be any useful function of the model histogram distribution of the parameter $d^n\theta$, of the curve segment of l_k crossing the region $r_{\setminus k}^{(a_j)}$.

The model driven information $p(a_j, v_{k,s})$ acts as a high-level constraint. It is sampled from the probability distributions of the model's structural description. The influence exerted by

this knowledge driven constraint is very useful in the solution search process. (i) It can stabilize the search space by continuously constraining it by the model driven information; (ii) It reduces the search space dimension and consequently introduces an upper bound to the combinatorial dimension of the search; (iii) Finally, it enhances the conditioning of the search space by better avoiding local optima. Moreover, this model driven constraint presents the intrinsic property of being expressed as a local measurement. This preserves the distributive form of the global criterion.

Finally, $p(l_{k,\delta s} | l_{k,s})$ of (4.5) can integrate different smoothing constraints. These constraints can be of invariant or of adaptive nature. $p(l_{k,\delta s} | l_{k,s})$ can be written as follows:

$$p(l_{k,\delta s} | l_{k,s}) = p\left(l_{k,s}, \frac{\partial l_{k,s}}{\partial s}, \frac{\partial^2 l_{k,s}}{\partial s^2}, \dots, \frac{\partial^D l_{k,s}}{\partial s^D}\right) \quad (4.13)$$

where the inner term of the RHS of (4.13) is the vector describing the truncated expansion of the local curve segment $l_{k,\delta s}$. D denotes the order of variational smoothing. Hence, the final form of (4.13) can assign a weighted penalty for each specific component of the expansion of $l_{k,\delta s}$. Since discontinuities play an important role in inverse visual problems, a less restrictive smoothing constraint, e.g., gradually adaptive smoothing, is useful for preserving discontinuities and precision of detection [38]. For invariant smoothing, (4.13) can be given by the same expression for all l_k and s . However, (4.13) can be intrinsically defined for each l_k according to *a priori* model driven or learned information.

4.7 Objective Function

The *log-pseudo-joint* solution can be obtained from (4.2) by optimizing the logarithm of (4.3) where we replace $p_{k,s}$ by (4.5), and $p(Y_{\delta_s} | r_{\setminus k, \delta_s}, l_{k, \delta_s})$, $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$, and $p(l_{k, \delta_s \setminus s} | l_{k, s})$ by (4.9), (4.12), and the desired smoothing functional (4.13), respectively. Note that

$\rho_{s,j} = \frac{|r_{\setminus k, \delta_s}^{(a_j)}|}{|\delta_s|}$. Hence, the resulting OF (*log-pseudo-joint*) is given by:

$$E = \sum_{k=1, \dots, K} \sum_{s \in L_k^I} \left(\sum_{j=1, \dots, m} \left(f_k \left(\bar{Y}_{\delta_s \text{ in}(k)}^{(a_j)}; \bar{Y}_{\delta_s \text{ out}(k)}^{(a_j)} \right) + g_k(a_j, v_{k,s}) \right) \frac{|r_{\setminus k, \delta_s}^{(a_j)}|}{|\delta_s|} \right) + h_k(l_{k, \delta_s}) \quad (4.14)$$

The three summations in (4.14) are for the K curves, for the lattice of each curve, and for the m local labels of $r_{\setminus k, \delta_s}$, respectively. The functions $\sum_{j=1, \dots, m} \rho_{s,j} \cdot f_k(\cdot; \cdot)$, $\sum_{j=1, \dots, m} \rho_{s,j} \cdot g_k(\cdot, \cdot)$

and $h_k(\cdot)$ are given by the logarithms of probabilities $p(Y_{\delta_s} | r_{\setminus k, \delta_s}, l_{k, \delta_s})$, $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$, and $p(l_{k, \delta_s \setminus s} | l_{k, s})$, respectively. f_k , g_k and h_k in (4.14) are generally weighted by three parameters $\alpha_{k,s}$, $\beta_{k,s}$ and $\gamma_{k,s}$, respectively. The weighting parameters, referred to as

“hyperparameters,” can be inferred from *a priori* knowledge, directly estimated from the data, or progressively computed from the processed solution. However, this stills an open question in computer vision research. OF (4.14) can be interpreted as an energy function totalizing the specific energies of the K DCs forming the model. Functions f_k and h_k are somehow similar to the local external and internal energies of a snake, respectively [33].

4.8 Implementation and Optimization Method

4.8.1 Object Model

The images are assumed to be the ST superposition of 2-D planar objects. Here, we assume that the intersections between any two different curves can only be disconnected points³⁰. Fig. 4.1 shows the data structure of the representation. In this instance, each node of the 2-D relational graph³¹ [3] $G_{O,\theta} = \{R_{O,\theta}, F_{O,\theta}\}$ of object O (O is projected onto the image plane from the pose position θ) represents a combination of some superposed planar subobjects. $R_{O,\theta} = \{R_{O,\theta,i}, i = 1, \dots, n_R\}$ and $F_{O,\theta} = \{F_{O,\theta,i}, i = 1, \dots, n_F\}$ are the sets of nodes and branches of $G_{O,\theta}$, respectively. Let $O = \{O_1, \dots, O_K\}$ be the set of the K planar subobjects. If each combination of the superposed subobjects is allowed to exist only once for one connected region in the projection, then the maximum number of 2-D region combinations, which is the maximum number of nodes in the 2-D graph, is $n_R = \sum_{i=1, \dots, K} K! / i! (K-i)!$. However, the number of nodes can be higher if we allow repeating labels for disjoint regions. Fig. 4.5 shows the 2-D relational graph of a specific 2-D model. Two different nodes, $R_{O,\theta,i}$ and $R_{O,\theta,j}$ of $G_{O,\theta}$, may eventually be connected by the branch $F_{O,\theta,i,j}$ if their corresponding regions α_i and α_j are adjacent and have therefore one common segment boundary of l_k of the planar subobject O_k . We note $\alpha_i = \alpha_j * in(k)$, where $in(k)$ denotes the inner region of the subobject

30

Intersection between two curves can not be continuous curve segments.

31

The relational graph of adjacent regions.

O_k , and $*$ denotes the superposition operation.

4.8.2 Explicit Forms of f_k , g_k and h_k

To write the explicit form of f_k , we recall the partition $r_{\setminus k, \delta s_{IO(k)}}^{(a_j)} = \left\{ r_{\setminus k, \delta s_{in(k)}}^{(a_j)}, r_{\setminus k, \delta s_{out(k)}}^{(a_j)} \right\}$ and

$p\left(Y_{\delta s} \mid r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right)$ in (4.8). $p\left(Y_{\delta s} \mid r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right)$ can be considered as the *sub-local likelihood*

function or the *sub-local* image formation process. Let $\xi\left(r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right)$ be the intensity

formation function transforming the *sub-local* template $r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}$ into the intensity 2-D

projected image. It can also be referred to as the ST projection function. Let $N(\mu, \sigma)$ be an

additive i.i.d. normal noise with μ mean and σ^2 variance. We assume that the local image

formation process can be written as:

$$Y_{\delta s}^{(a_j)} = \xi\left(r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right) + N(\mu, \sigma) \quad (4.15)$$

Hence, the *log-sub-local likelihood* can be written as [10]:

$$\begin{aligned} \ln p\left(N(\mu, \sigma) = Y_{\delta s}^{(a_j)} - \xi\left(r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right)\right) = \\ -\frac{|\delta s^{(a_j)}|}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{s \in \delta s^{(a_j)}} \left\| \mu - Y_s - \xi\left(r_{\setminus k, \delta s_{IO(k)}}^{(a_j)}\right) \right\|^2 \end{aligned} \quad (4.16)$$

(4.16) can be seen as a similarity measurement between Y and intensity template ξ . ξ can

be modeled as a sum of K specific intensity formation functions of the K curves. However, this kind of modeling seems inconvenient if we wish to deal with variable intensity objects over variable backgrounds [8], [1]. Additionally, estimating noise parameters is not a trivial task. Therefore, we propose to replace (4.16) by $\rho_{s,j} f_k \left(\bar{Y}_{\delta s_{in(k)}^{(a_j)}} ; \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right)$, where f_k is given

by the following relative measurement:

$$f_k \left(\bar{Y}_{\delta s_{in(k)}^{(a_j)}} ; \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right) = \text{sign} \left(\bar{Y}_{\delta s_{in(k)}^{(a_j)}} - \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right) \cdot \left\| \bar{Y}_{\delta s_{in(k)}^{(a_j)}} - \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right\|^\epsilon \quad (4.17)$$

$\rho_{s,j}$ is the same as discussed in Section 4.5.2, and henceforth equals to $\frac{|\delta s^{(a_j)}|}{|\delta s|}$.

Consequently, by replacing Y_s by the local mean $\bar{Y}_{\delta s_{(k)}^{(a_j)}}$ and the absolute measurement $Y_s - \xi_s$

by the relative measurement $\bar{Y}_{\delta s_{in(k)}^{(a_j)}} - \bar{Y}_{\delta s_{out(k)}^{(a_j)}}$ we avoid estimating noise parameters, modeling

curve templates and introducing an additional image intensity recovering process. The sign function in the RHS of (4.17) corresponds to the assumption that the inner region intensity of every subobject projection is greater than its outer region intensity. The distance term in (4.17) monotonically increases with the inner to outer intensity difference. For images with inverse intensities, we simply replace $\bar{Y}_{\delta s_{in(k)}^{(a_j)}} - \bar{Y}_{\delta s_{out(k)}^{(a_j)}}$ in (4.17) by $\bar{Y}_{\delta s_{out(k)}^{(a_j)}} - \bar{Y}_{\delta s_{in(k)}^{(a_j)}}$. For our

experiments, we set $\epsilon=1$. This gives the curve an energy proportional to the difference between inner and outer intensities. (4.17) is the local consistency measurement of the k th

curve matching with the image. It has the properties of precision and robustness due to the particular design of the local partitioning of $\delta s^{(a_j)}$ by $l_{k,\delta s}$ into inner and outer regions (see Fig. 4.2 and expression 4.8), and the averaging operation of Y on $\delta s_{in(k)}^{(a_j)}$ and $\delta s_{out(k)}^{(a_j)}$, respectively.

Function g_k ensures the local consistency of the k th DC with respect to the remaining curves using the *a priori* relational graph of the model. This consistency is carried out through algorithmic and functional constraints. As usual, the search process starts from an initial solution that is consistent with $G_{O,\theta}$. Algorithmic constraints are imposed during the sampling of the state space by forbidding the selection of states that modify $G_{O,\theta}$. g_k represents the functional constraint. In our implementation, we used a model-based statistical representation for each curve segment of l_k crossing a labeled region a_j . Thus, every branch $F_{O,\theta,(j,k)}$ of $G_{O,\theta}$ linking regions a_j and $a_{jk}=a_j * in(k)$ is augmented by a function of the statistical knowledge of the boundary local direction. This function is denoted by $g_k(a_j, \cdot)$ for the segment of the k th curve crossing region a_j . The abscissa scale τ of $g_k(a_j, \tau)$ is the local direction of the k th curve, where $0 \leq \tau \leq 2\pi$.

For the smoothing function h_k , we used a weighted ($\gamma_{k,s,1}$, $\gamma_{k,s,2}$, $\gamma_{k,s,3}$) sum of the curve length and the first and second quadratic variations of the curve local direction.

4.8.3 Optimization Method

OF (4.14) is obviously not convex. In fact, the convexity is directly related to the state space exploration process and thus to the definition of neighborhood state system. Stochastic or deterministic optimization methods can be used. Stochastic methods can generally converge toward a good estimate of the global optimal solution but they are computationally burdensome.

Deterministic methods can be either descent methods such as the steepest descent and the conjugate gradient methods or hill-climbing methods. Descent methods can get trapped in local optima. They are typically used for convex problems or when the initial state is close the global solution or when a locally optimal solution is sufficient. Deterministic hill-climbing methods require the definition of a neighborhood system, in the state space, used for exhaustively sampling the neighbors of the current state.

We here propose a descending search method with randomized sampling. Our model-based approach allows us to use deterministic search methods with some confidence in reaching an acceptable solution because the initial guess is generally close to the real solution, and the search strategy is continuously constrained by the model. Our method consists of continuously deforming the multiple dynamic curves in order to optimize energy function (4.14). Deformation is conducted iteratively by following the steps: (i) set the maximum deformation distance d ; (ii) randomly select a site $s \in L$, and define on L the neighboring disc $\Delta_{s,d}$ centered on s and with radius d ; (iii) for each curve C_k crossing $\Delta_{s,d}$, define the line segment $C_{k,\Delta_{s,d}} = C_k \cap \Delta_{s,d}$, erase $C_{k,\Delta_{s,d}}$ and then replace it by two

glued pieces of B-spline, Gaussian or straight line curve passing through s ; (iv) compute the energy variation using (4.14); (v) finally keep the new state if it is more likely than the previous one, otherwise restore the previous state. The parameter d is gradually decreased during the optimization process. Our search method can thus overcome many local optima using the randomized step-over deformation. Getting trapped in local minima can be further avoided by using block sampling (multiple curve samplers). In this way, a richer neighborhood system is created in the state space but for a much higher computational time.

4.9 Experimental Results

The approach was implemented as detailed in Section 4.8, using the data base shown in Fig. 4.1. We tested the method with three sets of synthetic images generated from specific initial images by adding an i.i.d. Gaussian noise of increasing standard deviation. Each initial image was formed by an additive ST projection of a simplified scene of planar objects. Images were all 128x128 pixels.

The same 2-D model was also used as initial state for all presented results. The corresponding relational graph shown in Fig. 4.5 can be replaced by its dual boundary graph. Fig. 4.5 (bottom-right) shows the grey level coding of the boundary segmentation of the model. Each boundary segment can be represented as a chain of geometric descriptors that can be compressed into the form of one statistical distribution for each geometric parameter. We choose only one descriptor that is the direction of the local tangent. Then, each $F_{O,\theta,i}$ of $G_{O,\theta}$ is augmented by the histogram distribution of its local tangent direction. This information

represents the branch attributes of $G_{O,\theta}$. However, directly sampling from this *a priori* information can create a model-biased behavior of the matching process, which essentially compromises the precision of the estimator. In order to overcome this limitation, we generated some trapezoidal window type curves from the thresholded histogram distributions. The relational graph was then augmented with this information. This can be interpreted as follows: all local deformations of a specific boundary segment receive the same *a priori* high-level influence if their geometric descriptors belong to the window opening of their corresponding graph branch. Fig. 4.6 shows the local tangent histogram distributions of four boundary segments of the model and their corresponding generated windows. The results are shown as labeled regions and curve segments, and are printed in coded grey levels.

The initial image of the first set is shown in panel $a0$ of Fig. 4.7. It was generated using a 2-D orthogonal projection of three overlapping planar objects on a non uniform background. Each object had a uniform attenuation of 1.0. Images $a1$ and $a2$ were generated from $a0$ by adding i.i.d. Gaussian noise with zero mean and a standard deviations of 1.0 and 3.0, respectively. The relational graph and its branch attributes were analytically inferred by segmenting and then labeling the model (Fig. 4.5). This *a priori* knowledge was used during the search process for the high-level constraining interaction. No global alignment was used. The parameter vector of the OF was set to $(\alpha_{k,s}, \beta_{k,s}, \gamma_{k,s,1}, \gamma_{k,s,2}, \gamma_{k,s,3}) = (1.0, 0.4, 0.05, 0.0, 0.7)$. $b0$, $b1$ and $b2$ are the results of matching with images $a0$, $a1$ and $a2$, after five full sweeps. One sweep is equivalent to 2^{12} deformation samples.

Fig. 4.8 shows the results obtained with the second set of noisy synthetic images. Model,

initial state, parameters, and number of iterations were the same as in Fig. 4.7. However, the uniform attenuation of the inner regions was replaced by a falling attenuation starting from the frontier. This can illustrate the X-ray projection of bone structures where the boundaries and the inner regions represent the cortical and spongy bone, respectively. The initial image a_0 (Fig. 4.8) was generated with the boundary and inner attenuations equal to 1.0 and 0.2, respectively. The boundary thickness varied between 2 and 3 percent of the image dimension. a_1 and a_2 were generated by adding an i.i.d. $N(\mu=0, \sigma)$ noise with $\sigma=1.0$ and 1.5, respectively. Negative intensities were set to zero. The results of matching with a_0 , a_1 and a_2 are presented in b_0 , b_1 and b_2 , respectively.

The initial image a_0 of the third set (Fig. 4.9) was generated using the parameters of the second set except the circular object was replaced by a square. Here, the difference between the model and the real solution is larger than in the previous experiments. The σ was equal to 0.8 and 1.2 for a_1 and a_2 , respectively. The matching results are shown in b_0 , b_1 and b_2 .

The results of matching the free model of multiple DCs with noisy synthetic images are very promising. The method showed a good robustness with respect to noise even with S/N of the order of 0.33 (Figures 4.7, 4.8 and 4.9), and to variable background (Fig. 4.7) and low contrast images (Figures 4.8 and 4.9). The boundary detection was precise and coherent. Curve crossing detection was also precise for different crossing angles even in severely degraded images. Unlike elastic or parametric models, our approach converge precisely to the intensity discontinuities even when the geometric shape of the real solution is not similar to the model (Fig. 4.9), and even with low contrast images (Figures 4.8 and 4.9).

Finally, we tested the importance of the high level influence in stabilizing the solution.

Therefore, we repeated some experiments with images from the second and third sets under the same conditions except β_k was set to zero. $b1$ and $b2$ in Fig. 4.10 represent the matching results with $a1$ of Fig. 4.8, and $a2$ of Fig. 4.9, respectively. Results proved the importance of the relational knowledge in stabilizing the search process, and thus in preserving the coherence of the solution by preventing unstructured convergence.

4.10 Discussion and Conclusion

In this paper, we presented a new probabilistic model-based vision system for segmentation and labeling of images of ST scenes. Model-based matching provides the advantage of carrying out simultaneously image segmentation and region labeling in one process. The wire frame model is made up of multiple superposed DCs. The OF was derived from the *pseudo-joint* estimator (4.2) [6], [5], [41] that has three important properties. First, the redundancy between the line fields was used to develop the data driven term offering a coherent and precise detection with respect to the observed image, and to incorporate local interaction between the solution and the model. Secondly, the local spatial dependency expressed over the observed image was used to develop a robust OF with respect to noise. Finally, the distributive property of criterion (4.2) allows us to express the global energy using exclusively local calculations which makes implementation possible on parallel machines. The resulting OF was made up of three parts which account for (i) the observed image, (ii) the model structural *a priori* knowledge, and (iii) the curve smoothing constraints.

The data driven term exhibits precision, local coherence and robustness behaviours in both

crossing and non crossing curves patterns, and offers a desired weak repulsive behaviour for the non-crossing curves. Unlike intensity-based methods, the data driven term was given by an inner-to-outer relative distance. This allowed us to avoid both the estimation of the noise parameters and the inconvenient modeling of the intensity template [1], [8]. In fact, our approach is essentially useful for images produced under varying conditions with non uniform backgrounds.

The model-based approach allowed us to introduce a high-level interaction in the global criterion. The relational knowledge of the model was integrated in the criterion using a line-region-graph hierarchical structure. Practically, the model driven term was derived from the *a priori* statistical distribution of the local parameters of the segmented multiple curve model. The statistical descriptions were then thresholded. This operation softened the model driven constraints and unlike the elastic and parametrical matching [2], [35], [30], it almost preserves the precision of the detection regardless the shape of the initial solution.

The resulting OF (4.14) was optimized using a descending method with relaxed randomized sampling. The start state was iteratively deformed in order to match the observed image. On the basis of the shown experiments, the search process was robust with respect to noise and offered good convergence to acceptable solutions.

The implementation can be enhanced by upgrading the geometric model to a volumetric representation that is certainly more realistic. Additionally, multiresolution techniques can improve the convergence by introducing long range interactions [37]. Moreover, stochastic modeling combined with an aspect graph [14] can be used to account for the imprecision in the pose estimate and the geometric fluctuations of the object. However, this expands the

state space dimension and makes stochastic search methods almost unavoidable.

Acknowledgements

H. Mallouche was supported in part by the National Sciences and Engineering Research Council of Canada (NSERC), the "Fonds pour la formation de chercheurs et l'aide à la recherche" of Quebec (FCAR), the Canadian Council of Professional Engineers-North American Life, and the cooperation France-Quebec, under scholarship numbers 167653, 961252, CCPE/NAL95-96, and 143690B, respectively. J.A. de Guise was supported in part by the NSERC of Canada. Y. Goussard was supported in part by the NSERC of Canada under Research Grant OGP0137417. We thank Dr Jérôme Idier of LSS, University Paris-Sud for valuable discussions.

References

- [1] Y. Amit, U. Grenander, and M. Piccioni, "Structural Image Restoration Through Deformable Templates," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 376-387, 1991.
- [2] R. Bajscy and S. Kovacic, "Multiresolution elastic matching," *Computer Vision, Graphics, and Image Processing*, vol. 46, pp. 1-21, 1989.
- [3] D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice-Hall, 1982.
- [4] J. Besag, "Towards Bayesian image analysis," *Journal of Applied Statistics*, vol. 16, no. 2, pp. 395-407, 1989.
- [5] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, pp. 179-195, 1975.
- [6] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 192-236, 1974.
- [7] R. A. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 285-348, 1981.
- [8] R. Brunelli and T. Poggio, "Template Matching: Matched Spatial Filters and beyond," *Artificial Intelligence Laboratory, MIT*, A.I. Memo no. 1549, 14 pages, 1995.
- [9] R. Curwen and A. Blake, "Dynamic Contours: Real-time Active Splines," in *Active Vision*, ed. A. Blake and A. Yuille, Cambridge: MIT Press, pp. 39-58, 1992.
- [10] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39-55, 1987.
- [11] M-P. Dubuisson and A.K. Jain, "Contour Extraction of Moving Objects in Complex Outdoor Scenes," *International Journal of Computer Vision*, vol. 14, pp. 83-105, 1995.
- [12] J.A. Feldman and Y. Yakimovsky, "Decision Theory and Artificial Intelligence: I. A Semantics-Based Region Analyser," *Artificial Intelligence*, vol. 5, pp. 349-371, 1974.
- [13] D. Geiger, A. Gupta, L.A. Costa, and J. Vlontzos, "Dynamic Programming for Detecting, Tracking, and Matching Deformable Contours," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 294-302, 1995.
- [14] U. Grenander and M.I. Miller, "Representations of Knowledge in Complex Systems,"

- Journal of the Royal Statistical Society, Series B*, vol. 56, no. 4, pp. 549-603, 1994.
- [15] W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, Cambridge: MIT Press, 1990.
 - [16] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour Models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.
 - [17] B.B. Kimia, A.R. Tannenbaum, and S.W. Zucker, "Shapes, Shocks, and Deformations I: The Components of Two-Dimensional Shape and the Reaction-Diffusion Space," *International Journal of Computer Vision*, vol. 15, pp. 189-224, 1995.
 - [18] K.F. Lai and R.T. Chin, "Deformable Contours: Modeling and extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1084-1090, 1995.
 - [19] D.G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, pp. 355-395, 1987.
 - [20] D.G. Lowe, *Perceptual Organization and Visual Recognition*, Boston: Kluwer Academic Publishers, 1985.
 - [21] H. Mallouche, J.A. de Guise, and Y. Goussard, "Probabilistic Model of Multiple Dynamic Curve Matching for A Semitransparent Scene," in *Vision Geometry IV*, R.A. Metler, A.Y. Wu, F.L. Bookstein, W.D.K. Green, Editors, Proc. SPIE 2573, July 1995, pp. 148-157.
 - [22] R. Malladi, J.A. Sethian, and B.C. Vemuri, "Shape Modeling with Front Propagation: A Level Set Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158-175, 1995.
 - [23] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W.H. Freeman, 1982.
 - [24] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76-89, 1987.
 - [25] J.D. McCafferty, *Human and Machine Vision: Computing Perceptual Organisation*, New York: Ellis Horwood, 1990.
 - [26] J.W. Modestino and J. Zhang, "A Markov Random Field Model-Based Approach to Image Interpretation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 606-615, 1992.
 - [27] R. Mohan and R. Nevatia, "Perceptual Organization for Scene Segmentation and Description," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 616-635, 1992.
 - [28] F. O'Sullivan and M. Qian, "A Regularized Contrast Statistic for Object Boundary Estimation-Implementation and Statistical Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 561-570, 1994.
 - [29] S. Sarkar and K.L. Boyer, "Perceptual Organization in Computer Vision: A Review and a Proposal for a Classificatory Structure," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 23, no. 2, pp. 382-399, 1993.
 - [30] L.H. Staib and J.S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061-1075, 1992.

- [31] G. Storvik, "A Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 976-986, 1994.
- [32] J.M. Tenenbaum and H.G. Barrow, "Experiments in Interpretation-Guided Segmentation," *Artificial Intelligence*, vol. 8, pp. 241-274, 1977.
- [33] D. Terzopoulos, "On matching deformable models to images," *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, D.C., pp. 160-163, 1987.
- [34] D. Terzopoulos, "Matching deformable models to images: Direct and Iterative Solutions," *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, D.C., pp. 164-167, 1987.
- [35] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, "Elastically Deformable Models," *Computer Graphics*, vol. 21, no. 4, pp. 205-214, 1987.
- [36] D. Terzopoulos, A. Witkin, and M. Kass, "Symmetry-Seeking Models and 3D Object Reconstruction," *International Journal of Computer Vision*, vol. 1, pp. 211-221, 1987.
- [37] D. Terzopoulos, "Image Analysis Using Multigrid Relaxation Methods," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 413-424, 1986.
- [38] D. Terzopoulos, "Regularization of Inverse Visual Problems Involving discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 129-139, 1986.
- [39] A. Witkin, D. Terzopoulos, and M. Kass, "Signal matching through scale space," *International Journal of Computer Vision*, vol. 1, pp. 133-144, 1987.
- [40] A. Witkin, J.M. Tenenbaum, "On the Role of Structure in Vision," in *Human and Machine Vision*, ed. J. Beck, B. Hope, and A. Rosenfeld, New York: Academic, pp. 481-543, 1983.
- [41] P. Whittle, "Stochastic Processes in Several Dimensions," *Bulletin of the International Statistical Institute*, vol. 40, pp. 974-994, 1963.

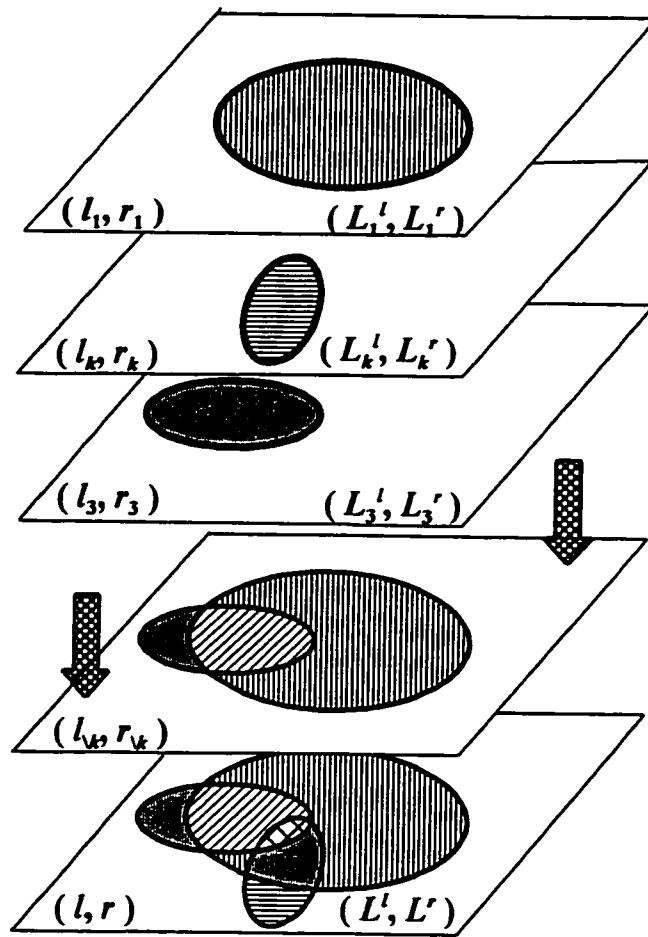


Fig. 4.1. Semitransparent 2-D projection of multiple contours. Each line l_k and its dual region r_k are defined on binary and discrete lattices L_k^l and L_k^r , respectively. The superpositions of the K curves and the K regions are denoted l and r defined on the K -ary lattices L^l and L^r , respectively. Here $K=3$.

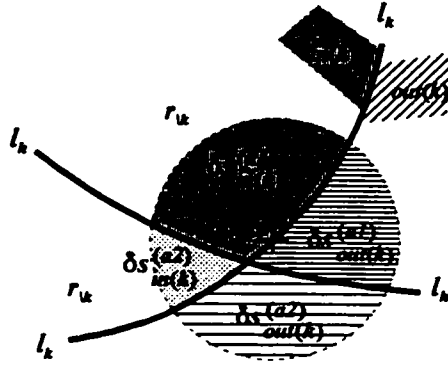


Fig. 4.2. The neighborhood system δ_s at site s of l_k is partitioned by segment l_{k,δ_s} into subsystems $\delta_{s_{in}(k)}$ and $\delta_{s_{out}(k)}$. Labels a_1 and a_2 are defined on the region process $r_{\setminus k}$.

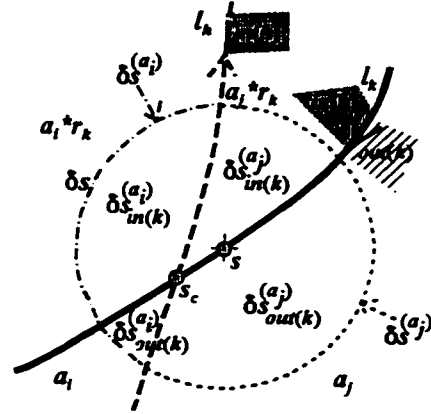


Fig. 4.3. Crossing curves: $s \in l_k$, δ_s is the neighborhood of s , and l_h crosses l_k . a_i and a_j are labels in process $r_{\setminus k}$.

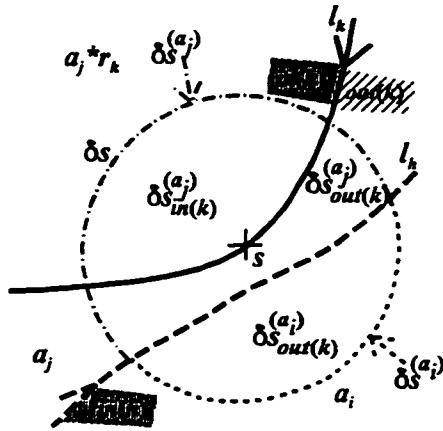


Fig. 4.4. Two non-crossing lines l_k and l_h .

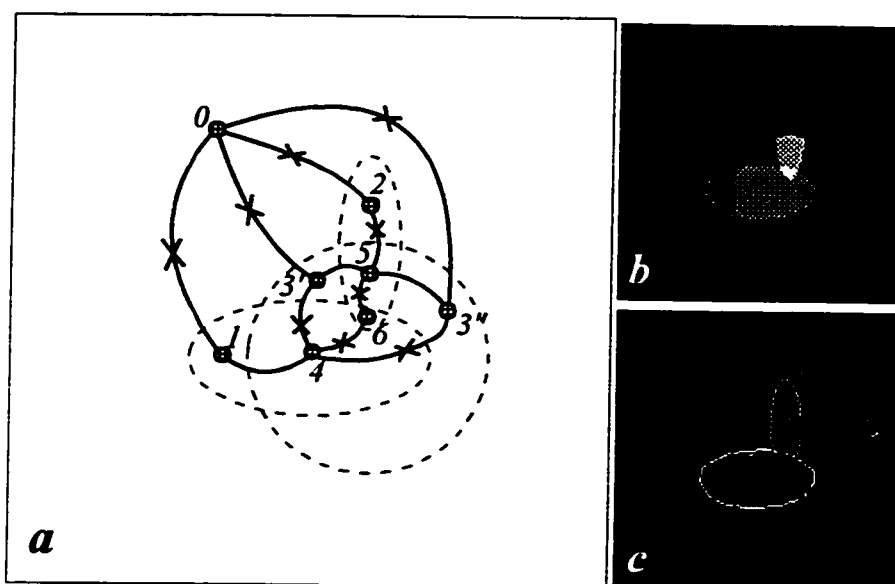


Fig. 4.5. *a*: The adjacency graph of the 2-D model. A node represents a connected region and a branch links two adjacent regions. The node 0 indicates the background, and nodes 3' and 3'' denote two disjoint regions having the same overlapping combination. *b* and *c*: The region and boundary segmentation (12 boundary labels) of the model coded in grey level.

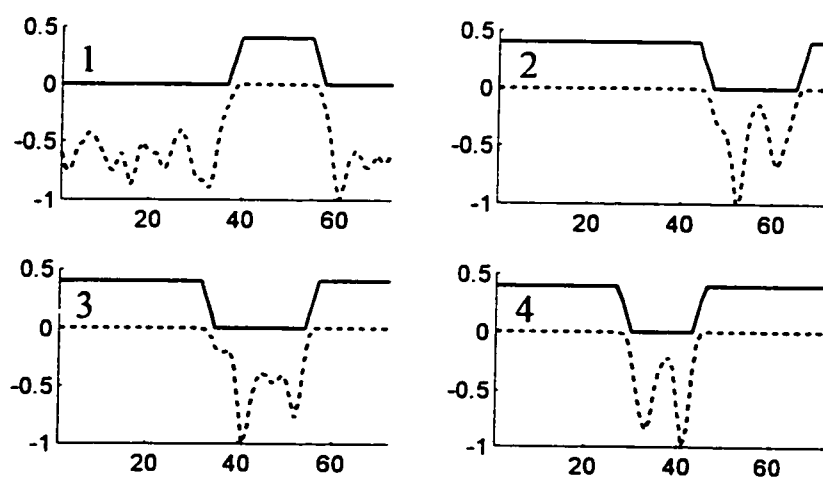


Fig. 4.6. The histogram distribution (dashed lines) of the local angular direction of the model's segments 1 to 4. The thresholded window type curves (solid lines) are generated from histograms. The abscissa represents the discretized local angular direction of 5° per step.

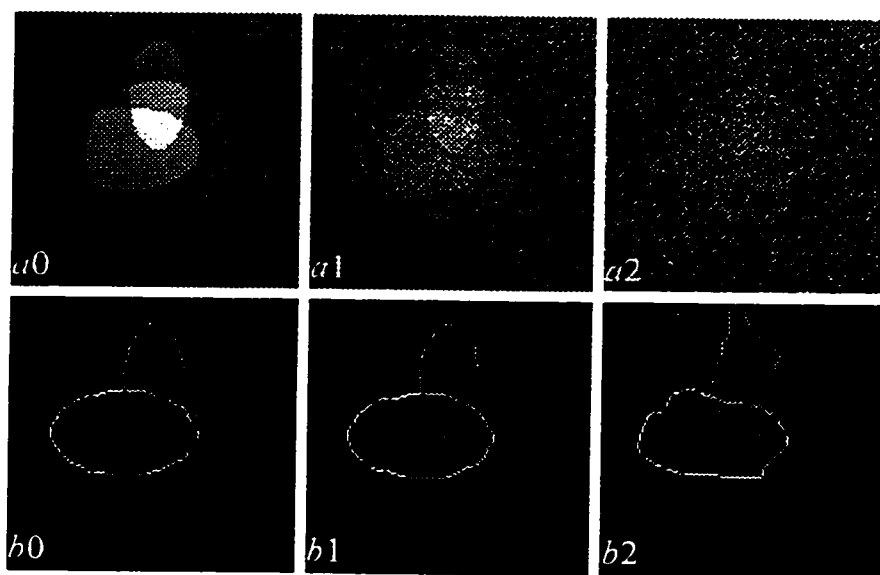


Fig. 4.7. a_0 : ST projection of three uniform objects over linearly variable background. a_1 and a_2 : i.i.d. $N(\mu=0; \sigma=1.0$ and $3.0)$ noise added to a_0 . b_0 , b_1 and b_2 : results of matching with a_0 , a_1 and a_2 .

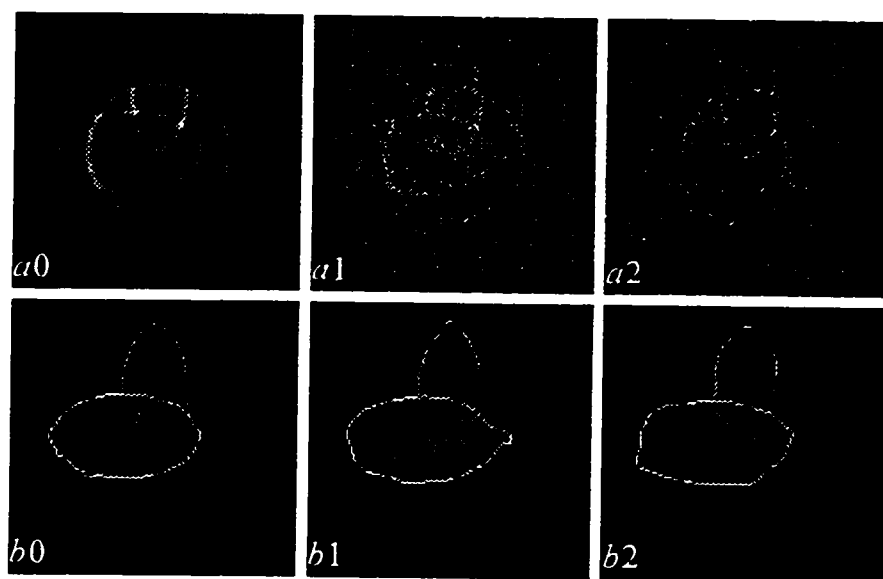


Fig. 4.8. a_1 : The subobject frontiers are set to 1.0 and the inner regions are set to 0.2. a_1 and a_2 : i.i.d. $N(\mu=0; \sigma=1.0$ and $1.5)$ noise added to a_0 . b_0 , b_1 and b_2 : results of matching with a_0 , a_1 and a_2 .

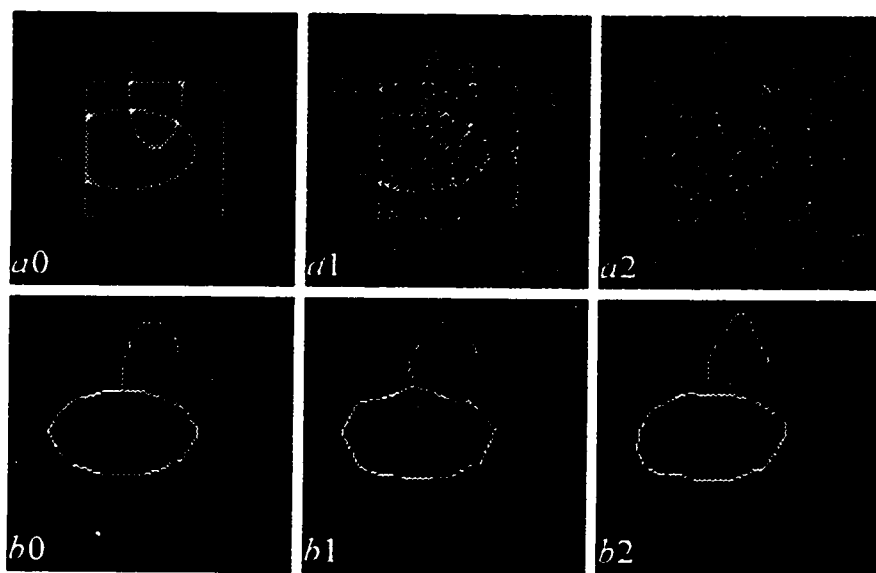


Fig. 4.9. $a1$ and $a2$ are generated from $a0$ by adding an i.i.d. $N(\mu=0; \sigma)$ noise of $\sigma=0.8$, and 1.2 , respectively. $b0$, $b1$ and $b2$: results of matching with $a0$, $a1$ and $a2$.

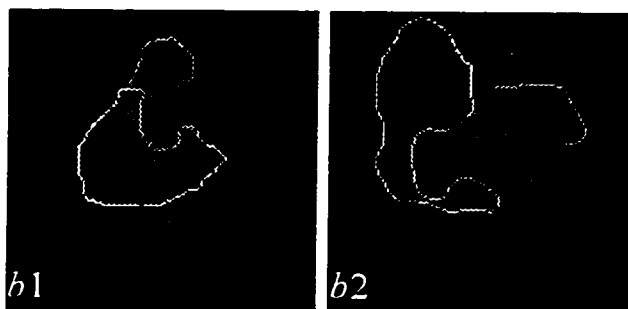


Fig. 4.10. Using the same parameters, model and start state, as before except β_k is set to zero, $b1$ and $b2$ are the results of matching with $a1$ of Fig. 4.8 and $a2$ of Fig. 4.9, respectively.

Chapitre 5

Multiples simples courbes fermées pour appariement de modèle libre avec images de scène semitransparente - La technique de déformation

Article: Multiple simple closed curve for free form model matching with images from semitransparent scene

L'article (Mallouche *et al.*, 1997c) du présent chapitre se situe dans le prolongement de l'article du chapitre 4. L'approche probabiliste est modifiée afin de représenter l'appariement avec un modèle plus riche et plus réaliste. L'article présente aussi la technique d'appariement par déformation brièvement décrite au chapitre 4.

5.0 Résumé

Cet article est divisé en deux parties. La première partie se concentre sur trois points: la modélisation volumétrique et l'aspect topologique des projections semitransparentes, la modification du modèle probabiliste pour intégrer un nouveau modèle 2-D, et les développements et discussions relevant du terme relié aux données et du terme relié au modèle. La deuxième partie est centrée autour de la technique d'appariement par

déformation. Trois points essentiels sont développés dans cette partie: la technique de déformation de la solution initiale; la formulation et l'implantation des contraintes algorithmiques; et le calcul de variation de la valeur du critère global (contrainte fonctionnelle).

Dans la première partie, on suppose qu'un objet 3-D peut être représenté par un assemblage de sous-objets simples de forme libre planaire, conique ou cylindrique. La projection 2-D d'un sous-objet peut être représentée par une courbe dynamique extrémale combinée ou non avec une ou deux courbes dynamiques internes. Une courbe dynamique est définie dans un cadre topologique comme équivalente à une simple courbe fermée (Engelking et Sieklucki, 1992). La projection 2-D d'un objet est la superposition des projections de tous les sous-objets qui le composent. Le critère de MPPC est par la suite ajusté par rapport à cette nouvelle représentation du MMC incluant des courbes internes. Le terme relié aux données est aussi ajusté. Le terme relié au modèle est ensuite discuté en détail. On exprime ce terme comme une mesure de similarité entre les segments de courbes du modèle et les segments de courbes de la solution. En tenant compte de la nature déformable d'objets et de l'importance de la précision de détection, les paramètres locaux proposés pour la mesure de similarité sont la direction locale et ses dérivées curvilignes. En tenant compte des variations géométriques de l'objet, on calcule les histogrammes des paramètres locaux pour chaque segment de courbe du modèle. Afin de préserver la précision de la détection, les histogrammes sont seuillés en fonctions pseudo-binaires. Le terme relié au modèle est donné par l'ensemble de ces fonctions.

La deuxième partie de l'article est consacrée à la présentation de la méthode

d'appariement. Celle-ci est basée sur une déformation aléatoire et progressive de la solution jusqu'à l'atteinte d'une solution optimale. Une courbe l_k et un sommet s_a sont choisis aléatoirement pour chaque tentative de déformation. Le segment à déformer $l_{k,(s_0,s_1)}$ est ensuite délimité. s_0 et s_1 désignent les extrémités du segment. $l_{k,(s_0,s_1)}$ est remplacé par deux segments liés de splines cubiques notés $l_{k,(s_0,s_a,s_1)}$. La génération du nouveau segment intègre les contraintes de la préservation de SCF.

Pour être accepté, l'échantillon généré doit aussi vérifier les contraintes algorithmiques et la contrainte fonctionnelle. Les contraintes algorithmiques assurent la préservation du contenu relationnel de la solution initiale tout au long du processus de déformation. On démontre que le nouvel échantillon doit répondre à une condition de préservation de l'étiquetage séquentiel de $l_{k,(s_0,s_1)}$ et à une condition d'inclusion de l'ensemble d'étiquettes de région affectée par la déformation dans l'ensemble des étiquettes de $l_{k,(s_0,s_1)}$. De son côté, la contrainte fonctionnelle traduit le choix de la méthode d'optimisation descendante. Pour chaque nouvel échantillon (une déformation), on calcule la variation dans la valeur de la fonction à optimiser. La déformation est acceptée si elle génère un état plus probable que l'état précédent.

Finalement, l'approche est expérimentée avec succès sur des images synthétiques fortement bruitées puis avec des images rayon-X d'un objet synthétique de géométrie complexe.

Les contributions de l'article sont les suivantes:

1. formulation du modèle à multiples SCF dans un cadre topologique;
2. représentation plus réaliste des projections 2-D des objets volumétriques à l'aide des SCF extrémales et internes;
3. démonstration de la proposition de la préservation du graphe relationnel;
4. développement de la nouvelle méthode descendante de recherche par déformation aléatoire et progressive de solution.

**Multiple Simple Closed Curves for Free Form Model Matching
with Images from Semitransparent Scene**

Hail Mallouche, Yves Goussard, Jacques A. de Guise

Correspondence should be sent to:

Hail Mallouche

Biomedical Engineering Institute

Ecole Polytechnique of Montreal

P.O.Box 6079, Station "Centre-Ville"

Montreal, Quebec, Canada

H3C 3A7

Phone: (514) 340-4198

Fax: (514) 340-4611

Email: hail@grbb.polymtl.ca

Multiple Simple Closed Curves for Free Form Model Matching with Images from Semitransparent Scene

Hail Mallouche, Yves Goussard, Jacques A. de Guise

Abstract. Segmentation and labeling of X-ray medical images are untrivial tasks. These images are complex with low contrast and low signal to noise ratio (S/N). Hence, edge detectors produce oversegmented elements. Grouping and organization of these elements is not straightforward especially for deformable objects. Elastic models have been successfully used for matching slightly deformable objects. Their performance and precision can be heavily compromised when the disparity between the model and the image is large and when the image contrast is weak and variable.

In this paper, we present a new approach to matching free form deformable models with X-ray images. Unlike elastic models, free forms (or wire-frame models) do not bias the boundary detection toward the initial state. We assume that objects can be assembled using a set of elementary 3-D free-forms (*subobjects*): planar, conic and cylindric. Multiple deformable curves (*simple closed curves*) are used to model the 2-D semitransparent projection of the subobjects. Extremal curves and inner curves are gathered in a topological scheme. We develop a probabilistic criterion for matching the 2-D projection of the 3-D model with the observed image. The criterion is made up of three terms. (i) The *data driven* term uses the *line-region* duality of *simple closed curves* to create a natural and coherent interaction between the multiple curve solution and the image. (ii) The *model driven* term

carefully incorporates the statistical representation of the structural knowledge of the 2-D model. (iii) The *smoothing term* helps stabilize curve matching when the images are noisy. The resulting objective function is optimized using an iterative descent technique that randomly deforms the solution in order to match the observed image. The deformation process has to preserve the *simple closed* curve representation and the *relational graph* of the initial model. We show that the latter condition is equivalent to two algorithmic constraints: the line labeling preservation and the region labeling inclusion. Finally, we present successful experimental results of matching wire-frame models with noisy synthetic images and X-ray images of a synthetic object.

Index Terms. Deformable simple closed curves, wire-frame models, free form matching, semitransparent scene analysis, segmentation and labeling of X-ray images, functional and algorithmic constraints.

5.1 Introduction

Image analysis is of great interest in medical imaging. The combination of medical imaging and computer vision allows us to introduce new techniques for automating medical procedures, improve reconstruction modalities and propose new interventional strategies. Atlas matching, invasive or non invasive surgical planning and intervention, and 3-D reconstruction from a small number of views are just examples of applications that directly benefit from this field of research.

The goal of this work is to develop a model-based computer vision system for segmentation, labeling and recognition from radiographic images. The ultimate goal is to apply the system to X-ray images of complex scenes such as the spinal bone structure. Besides the low intensity contrast and the low S/N of X-ray images, the semitransparent superposition of the structures renders this task very difficult. Moreover, the variable (in, e.g., scale, position, elongation, etc.) and deformable geometry of human organs increases the complexity of the images.

Many computer vision techniques have been applied to medical images from different modalities. Low level boundary detection and surface computation techniques were applied to CT, ultrasound or X-ray images to extract different anatomical parameters (Kass *et al.* 1988; Ayache *et al.* 1992). Other techniques were developed to infer the axes of symmetry of elongated organs. These axes were then used in orthographic or stereographic matching for 3-D inference (Close 1991; Suetens *et al.* 1988). Elastic models have been used for atlas matching (Bajscy and Kovacic 1989). Dynamic curves have also been used for boundary extraction (Storvik 1994). However, all these techniques have been restricted to the use of one curve or several disjoint curves for tomographic or simple X-ray images.

Our approach is mainly based on incorporating free form multiple curve representations into structural framework. Unlike elastic models, free forms are material free, can move and be deformed in all directions without being compromised by the initial unstressed state of the model (Terzopoulos *et al.* 1987). Additionally, this choice is justified by our need for a precise detection of the intensity discontinuities and to delineate objects with highly variable geometric forms. Structural information is important for recognition in biological vision

(Tenenbaum and Barrow 1977), and we assume that complex objects can be assembled using a set of elementary subobjects and conformably to a specific 3-D relational graph. We propose three simple free forms referred to as subobjects: planar, conic and cylindric. Then, we assume that the viewing position is approximately known. This allows us to produce a 2-D model for any semitransparent projection. The 2-D model is made up of multiple superposition of two kinds of simple closed curves: extremal and inner. Our approach to segmentation and labeling consists of matching the 2-D free form model with the observed image.

Matching is performed in a stochastic framework under the assumptions of local Markovian property of the model, the local dependency of the observed image, and by taking advantage of the line-region duality of simple closed curves. The objective function (OF) is made up of three terms. (i) The *data driven* term that measures the consistency between the solution and the observed data. It is developed to ensure the precision and the coherence of the discontinuity detection. (ii) The *model driven* term that constrains the solution using unbiased relational *a priori* knowledge inferred from the statistical representation of the 2-D model structure. (iii) The *smoothing term* that introduces some desired adjustable rigidity to the deformable curves.

Finally, the OF is optimized using a descent search method with randomized sampling based on an iterative deformation of the solution. This technique introduces long range interactions that create rich transition possibilities. Such search techniques can overcome many local optima. The deformation technique is constrained to preserve the initial 2-D relational graph of the model. For this purpose, any proposed deformation must fulfill two

algorithmic conditions. Experiments with synthetic images and X-ray images illustrate the efficiency, robustness and precision of the proposed approach.

The paper is organized as follows. The next section is devoted to the problem statement. We discuss the related works in Section 5.3. Our approach is globally presented in Section 5.4. Details about our modeling technique for images of semitransparent scenes using simple closed curves are given in Section 5.5. We also include a topological foundation introducing the notions of extremal curve and inner curve. In Section 5.6, we develop the global probabilistic criterion, we present and discuss the computational issues of the *data driven* term and the statistical representation of the *model driven* term. Section 5.7 describes the search method, the deformation technique and the algorithmic and functional constraints. In Section 5.8, we present the experimental results. Finally, the approach and future research are discussed in Section 5.9.

5.2 Problem statement

Recognition of deformable objects is still an open problem in computer vision. Generally, recognition consists of identifying the object and the viewing position or the pose (translation, rotation and scale) (Grimson 1990; Lowe 1985). Different approaches have been proposed for object recognition. The techniques are generally centered on (i) suggesting a hypothetical object, and (ii) iterating between estimation of the pose and search for additional evidence by model matching. These matching methods perform well with rigid or slightly deformable objects (Grimson 1990; Lowe 1985).

In controlled scenes, e.g., medical scenes, we can assume that the object is known and that the viewing position can be approximately estimated by several means. In fact, the goal of the present work is to develop a matching process between X-ray images of complex non rigid objects and an *a priori* known 2-D model. Note that the 2-D model can be directly induced by projecting the 3-D model from the estimated pose. The matching process aims at recognizing and labeling the constituent parts of the object, and at precisely delineating the intensity discontinuities of the object.

The process has to overcome many difficulties that originate from: (i) The basic nature of the scene: Human organs have a deformable and complex geometry and are arranged densely. In addition, the organs are made up of different types of tissues that each has its own X-ray attenuation. (ii) The image formation process: Both the X-ray penetration capacity and the tissue attenuation property with respect to X-ray create the semitransparency of the image formation process referred to as semitransparent scene. Consequently, resultant images are complex and are superimposed on a variable background. In addition, X-ray images have a weak S/N due to photon diffusion within the material. (iii) The ill-posed nature of the inference problem and the structure of the search space: The ill-posedness is more severe due to the first and second difficulties. The complexity and the weak S/N reduce the visibility and the observed evidence. In this instance, any edge detector provides an over-fragmented set of edge elements. This can drastically increase the dimension of the search space. In fact, the search space dimension can be reduced by using geometric constraints that naturally reflect the perceptual organization criteria. However, these constraints are not always suitable for non rigid object matching. Moreover, the choice of the search method is not trivial especially

for non convex problems. Generally, the choice of the search method is a tradeoff between the computational cost and the ability to converge toward an acceptable solution.

In order to overcome the above-mentioned difficulties, we propose to perform detection and segmentation in complex X-ray images using a model-based approach guided by a single objective function. Specific questions can emerge from this scheme: (i) What is the adequate model that should be used? (ii) How can we develop the objective criterion in a coherent and harmonious manner with the selected model? (iii) What kind of additional model-based knowledge can be used to constrain the search without compromising the precision of the boundary detection? (iv) What is the most convenient search method? Answers to these questions are provided in the sequel.

5.3 Related work

Model-based matching methods for recognition, segmentation and labeling can be gathered with respect to (i) matching in the feature correspondence space or (ii) matching with nonrigid models.

On one hand, recognition can be seen as a search process in the correspondence space. One node in the correspondence space represents a set of matching pairs between an image feature and a model feature (Grimson 1990; Horaud and Skordas 1989). Search techniques in the correspondence space, e.g., Hough transformation (Ballard and Brown 1982), search in the interpretation tree (Grimson 1990; Lowe 1985; 1987) and graph matching techniques (Horaud and Skordas 1989), have essentially been used for matching with rigid objects.

Consequently, these techniques are generally based on global hypotheses locally verified by feature matching. These methods behave well with rigid and polyhedral objects. However, they encounter serious difficulties with non rigid objects because that the extracted features do not always verify the pose invariant condition, and that the feature similarity-based matching cannot account for large disparities of highly deformable objects (objects in medical scenes).

On the other hand, both dynamic contours and elastic models of non rigid and non polyhedral objects have been used in parametric and non parametric representations for matching simple images of opaque scenes or tomographic medical images (Terzopoulos 1987; Terzopoulos *et al.* 1987; Staib and Duncan 1992; Kass *et al.* 1988). Elastic methods can be described as follows: (i) The model is deformed by the external forces derived from the observed image. (ii) The solution is an equilibrium between the external forces and the inner elastic forces related to the deviation of the solution from the initial state. Hence, the model recovers its initial state when external forces vanish. These methods have two major drawbacks (Bajscy and Kovacic 1989): (i) The convergence is sensitive to the initial state. (ii) The precision of detection is biased by the initial state. This is due to the elastic property of the model that is generally inconsistent with the reality.

Dynamic contours have mainly been used as free form (non parametric and material free) and represented by snakes (Terzopoulos 1987). Generally, snake-image matching is carried out by optimizing an energy function made up of an inner energy (or smoothing term) and an external energy measuring the goodness of fit between the image and the snake. Free-forms have been used for intensity discontinuity detection (Ayache *et al.* 1992; Storvik 1994).

Unlike elastic models, these forms are not constrained by their initial state. They can be deformed in all directions in order to match with intensity discontinuities. However, application of free-forms has been limited to one curve or several disjoint curves.

Parametric representations have also been used for deformable models (Staib and Duncan 1992; Lai and Chin 1995). They are applicable when objects can be parametrically described as $F(x, a) = 0$. x is a 2-D or a 3-D coordinate vector and a is the parameter vector. The major advantage of this representation is the compactness of the search space. In a stochastic framework, a is represented by its statistical distribution $p(a)$ (Staib and Duncan 1992). Then, the solution results from tradeoff between the observed image and the *a priori* distribution $p(a)$ and therefore suffers from the same drawbacks as the elastic models. Mainly, the precision of detection decreases when analysed images are noisy with a variable spatial contrast, or when the model is largely different from the real object.

Unlike the search in the correspondence space, methods of deformable templates generally use a single objective function. The matching is achieved by iteratively optimizing this function using either deterministic or stochastic methods. Deterministic methods as, e.g., variational methods mostly used to solve the dynamic equation of elastic models (Terzopoulos 1987; Bajscy and Kovacic 1989) can be trapped in local optima due to their limited transition neighborhood that restricts their usefulness to convex or nearly convex problems. Stochastic methods generally have a better assurance of converge than deterministic ones but they are computationally burdensome (Storvik 1994).

In this paper, we develop a new method for segmentation, labeling and detection of X-ray images. In order to reduce the dimension of the search space, our approach is based on direct

matching between a known deformable model and the observed image. The model is formed of the superposition of multiple free dynamic curves. This accounts for the semitransparency of X-ray images. Our work is related to the free form methods that have a higher precision of detection than elastic and parametric methods. However, we use a topological framework in order to model semitransparent projection of simple volumetric subobjects, ensure cross-curve coherence and constrain the matching procedure by *a priori* model-based structural information. Our work is also related to elastic models in the sense of preserving the structure of the solution.

5.4 Method and assumptions

We now present the general description and the basic guidelines of our approach. The system uses a model-based approach to achieve matching between a 2-D model and 2-D images of semitransparent scenes. The approach is made up of two levels: (i) the geometric modeling of the objects and (ii) the model matching with the observed image.

In order to select a geometric model, we have to balance its conciseness and its descriptive power (Ballard and Brown 1982), and we have to consider the complex nature of the objects to model. Many 2-D or 3-D geometric representations have been used for object modeling in computer vision (Besl 1988; Ballard and Brown 1982). 3-D shapes can be described either with volumetric models or with bounding surfaces. The former is useful when the inner information about the objects has to be inferred. Otherwise, surface (boundary) modeling is adequate and offers a compact representation. Surface representation

can be parametric (generalized cylinders, superquadrics, hyperquadrics, geons) or any blended function of these forms (Cohen and Cohen 1996). These shapes are useful for modeling symmetric objects. However, they lack generality for representing a large class of highly deformable objects. Non parametric surface representations can also be used. They are made up of a set of connected surface patches or a free form surface.

In order to reduce the complexity of the geometric models, the structural representation has been extensively used in computer vision for articulated and complex objects (Marr 1982; Pentland and Horowitz 1991; Ballard and Brown 1982). It allows us to model a family of objects using structural assembling of subobjects that generally belong to a set of simple geometric forms. The assemblage complies with the relational graph describing the intrinsic connectivity of the object. Modeling can be either 2-D or 3-D depending on the application and the objects to represent. 3-D modeling of polyhedral objects has been successfully used for direct matching with 2-D images of opaque scenes (Lowe 85; 87). However, matching is essentially realized using the estimated 2-D projection of the 3-D model. Our approach uses a similar representation. Here, our goal is to design a model that is: (i) capable of representing human organs and particularly bone structures (non rigid nature of the objects); (ii) simple; (iii) and consistent with the information content of the images (dominant boundary information) and the image formation process (semitransparency).

Consequently, we choose a free deformable 2-D model of multiple simple closed curves. The boundary representation is convenient for X-ray medical images because the filling tissues of the organs generally have variable photon attenuation and because organs are projected onto variable background. In addition, a free from model allows us to account for the large

deformable nature of the objects to detect. 2-D modeling using a multiple closed curve representation is also simple. It is interesting due to the continuity and closure preservation of simple closed curves that implicitly overcomes the edge connectivity and the region merging difficulties in perceptual organization and hence reduces the search space dimension. The superposition of multiple curves represents the semitransparent overlapping of the structure in X-ray images. The model can also represent the 2-D projection of 3-D structural geometry that we consider convenient to represent the complex geometry of the human organs.

Regarding model matching with the observed image, it can be seen as a segmentation and labeling operation. For images from controllable scenes such as medical X-ray images, we can conveniently assume that the identity of the object to detect is known. Hence, we assume that we dispose of *a priori* deformable model representing the object. In fact, model matching allows us to label and delineate the constituting parts (subobjects) of the object. Matching is performed in the 2-D image plane between the semitransparent projection of the model and the observed image. For this, we assume that an estimate of the viewing position is available. Matching is carried out along the following lines: (i) Deform the initial solution in order to match with the image; (ii) Use intrinsic geometric constraints and knowledge-based constraints in order to stabilize the search space and reduce its dimension; (iii) Use a descent search technique with a dense and long-range transition function to reduce the computational cost and increase the local convexity of the search space; (iv) Finally, use a stochastic framework to develop a single distributive objective function that measures global consistency between the solution, the image and the model using local interactions. This also allows us

to optimize the objective function using local deformations and calculations.

5.5 Object modeling

Here, we assume that each object can be constructed from simple subobjects using an intrinsic 3-D relational graph (Ballard and Brown 1982; Marr 1982). For our specific application, we limit the types of subobjects to three free generalized shapes (Fig. 5.1): cylindric, conic, and planar. The 2-D semitransparent projections of the cylindric, conic and planar shapes can be described using (i) one extremal curve and two inner curves, (ii) one extremal curve and one inner curve, and (iii) one extremal curve only. The extremal curve is the silhouette that delineates the outer limits of the 2-D semitransparent projection of the subobject. The inner curve represents an intensity discontinuity in the observed image resulting from a 3-D surface discontinuity of the subobject. We assume that both extremal and inner curves are simple closed curves. In the next sections, we give the precise definitions of extremal and inner curves.

The 2-D semitransparent projection of the assembled object from a known viewing position can be described by the superposition of the 2-D projection of the constituting subobjects. The resulting superposition defines a 2-D relational graph that can be seen as the projection of the 3-D relational graph from a known viewing position. The 2-D graph is a vertex in the aspect graph of the model (Dickinson *et al.* 1992).

We now give the definition of *simple closed* curves with the line region duality before introducing the relations between extremal and inner curves. These foundations are essential

to formulating the probabilistic model and the deformation technique.

5.5.1 Simple closed curve - Foundations

Definition³²: A *simple closed curve* (SCC) is a continuous closed curve that fulfills the following conditions: (i) The curve has no multiple points. In other words, it does not cut itself; (ii) The curve has a unique local orientation.

Definition: A set of points S is said to be *connected* if any two of its points can be joined by a SCC defined on S .

Curve-region duality³³: The complement of a SCC l in the image plan r consists of two disjoint regions r^I and r^O referred to as the inner region and the outer region, respectively. Each of r^I and r^O is a connected domain.

Definition: A domain S is *simply* connected if the inner region of any SCC within the domain is also part of this domain. When *disjoint* holes exist within the domain S , we say that S is a *multiply* connected domain. A domain with one hole is said *doubly* connected (Fig. 5.2). The connectivity of a domain with $n-1$ holes is n , where the domain is called *n-tuply* connected.

Let $z(s)=x(s)+i.y(s)$ be the parametric form of l . If we assume that r^I lies to the right of l then for each point $z(s) \in l$ the normal to l given by $n(s)=-y'(s)+i.x''(s)$ is directed

³² By analogy with Jordan curve (Engelking and Sieklucki 1992; Cohn 1967)

³³ By analogy with *Jordan curve theorem*, 1887 (Engelking and Sieklucki 1992; Cohn 1967).

into r^I . This means that the point $Z(\eta) = z(s) + \eta \cdot n(s)$ lies into r^I for any sufficiently small $\eta < 0$. Equivalently, we write that r^I lies to the right of l if for every $s \in l$, a real number $\Delta(s) > 0$ exists such that: if $0 < \eta < \Delta(s)$ then $Z(-\eta) \in r^I$ and $Z(\eta) \in r^O$. Hence, the right-hand normal is called the *inner* normal to l with respect to r^I . If l has clockwise *CW* (counter clockwise *CCW*) orientation, then r^I and r^O are also called the right-hand (left-hand) side and the left-hand (right-hand) side of l , respectively.

5.5.2 Extremal and inner curves

Now, we define extremal and inner curves. Let r be the connected set (region) of sites defined by the SCC l . The inner region r^I of r can be defined as $r^I = in(r) = \{s \in r \mid \exists \eta_s > 0, B^2(s, \eta_s) \subset r\}$ and is equivalent to the largest *open* subset defined on r . We can write $r^I = r \setminus l$. r is the set of all convergence points of r^I and is referred to as the *closure* set of r^I .

Now, let l_k and l_{k1} be two *simple closed* curves (SCCs). We say that l_{k1} is an *inner* curve of l_k if (i) $l_{k1} \subset r_k$ and (ii) l_{k1} can be partitioned as $l_{k1} = \{l_{k1}^{\setminus k}, l_{k1}^k\}$ where $l_{k1}^{\setminus k} \neq \emptyset$ and $l_{k1}^k \neq \emptyset$ are two piecewise curve segments given by $l_{k1}^{\setminus k} = l_{k1} \cap r_k^I$ and $l_{k1}^k = l_{k1} \cap l_k$, respectively. l_k is then referred to as the *extremal* curve containing the inner curve l_{k1} . l_k can be

partitioned as $l_k = \{l_k^{\setminus k1}, l_k^{k1}\}$ where $l_k^{\setminus k1} = l_k \setminus l_k^{k1}$ and $l_k^{k1} \equiv l_{k1}^k$. l_k^k and l_k^{k1} are referred to as *double* segments as opposed to the *simple* segments $l_{k1}^{\setminus k}$ and $l_k^{\setminus k1}$ (Fig. 5.2). Consequently, the pair $\{l_k, l_{k1}\}$ defines a special case of a *doubly* connected domain where the intersection between l_k and l_{k1} is not empty and generates one *doubly* continuous segment.

From a general viewing position, we assume that the semitransparent projection of a free conic shape can be modeled by a pair $\{l_k, l_{k1}\}$ of extremal and inner curves. Additionally, we assume that the 2-D projection of the free cylindric shape can be modeled by $\{l_k, l_{k1}, l_{k2}\}$, where l_k , l_{k1} and l_{k2} are an extremal curve and two disjoint inner curves, respectively. Thus, $\{l_k, l_{k1}, l_{k2}\}$ is a special case of a 3-tuply connected domain where the intersection between l_k and each of l_{k1} and l_{k2} generates only one continuous *doubly* segment.

5.6 Probabilistic model

In this section, we briefly present the probabilistic criterion and its final expression. The search process for matching the multiple SCCs $l = \{l_k, k = 1, \dots, K\}$ with the observed image Y is guided by the optimization of the global criterion. We refer the reader to (Mallouche *et al.* 1995; 1997a) for the theoretical foundation of the criterion. Let $E(l, Y, M)$ denote the objective function where M is the *a priori* known 2-D model. M is a set of K SCCs denoted $l^M = \{l_k^M, k = 1, \dots, K\}$. Let l^0 denote the initial state that is usually chosen to be identical to

I^M .

We wish $E(I, Y, I^M)$ to exhibit the following convenient properties: (i) distributive form of the global criterion; (ii) consistency and coherence between the SCCs; (iii) consistency, robustness and precision of the resulting solution with respect to observed image Y .

We use a stochastic framework to develop the global criterion. This allows us to benefit from interesting advantages. (i) We can make use of *a priori* knowledge about the object. (ii) We can introduce specific assumptions to build the final criterion. (iii) Finally, the behavior of the method can be directly analyzed with respect to the criterion properties and the proposed assumptions.

5.6.1 Global criterion

Thus far, many probabilistic estimators have been used for image restoration, analysis and understanding (Besag 1989; Derin and Elliot 1987; Marroquin *et al.* 1987). The *maximum a posteriori* (MAP) criterion, $\max_I p(I|Y)$, has been widely used because it expresses the

global scene understanding and the reconstruction problem in a single objective function.

Often, $p(I|Y)$ is of complex nature and of very high dimensionality. We argue in (Mallouche *et al.* 1997a) that the *maximum pseudo-joint* (MPJ) solution can be considered as a reasonable approximation to the *maximum joint* solution, $\max_I p(I, Y)$, and consequently to

the MAP solution when adequate Markovian and local dependency properties are assumed

for the model and the image, respectively. Let s be a site defined on a lattice L . δs is a neighborhood system defined in L and centered on s . Let $|\cdot|$ denotes the cardinal operator of a set or subset. The MPJ solution is given by maximizing the *pseudo-joint* distribution:

$$\Phi = \prod_s p(l_{\delta s}, Y_{\delta s}) \quad (5.1)$$

where $p(l_{\delta s}, Y_{\delta s})$ is the local joint probability distribution. Let $l_{k,s}$ be a site s of the SCC l_k , $l_{k,\delta s} = \{l_{k,t} | t \in \delta s\}$ and $l_{\setminus k,\delta s} = \{l_{1,\delta s}, \dots, l_{k-1,\delta s}, l_{k+1,\delta s}, \dots, l_{K,\delta s}\}$. Φ can then be written as follows:

$$\Phi = \prod_{k=1,\dots,K} \prod_{s \in l_k} p(l_{k,\delta s}, l_{\setminus k,\delta s}, Y_{\delta s}) \quad (5.2)$$

Using the line-region duality we can substitute $r_{\setminus k,\delta s}$ for $l_{\setminus k,\delta s}$ in Φ , where $r_{\setminus k,\delta s} = \{r_{1,\delta s}, \dots, r_{k-1,\delta s}, r_{k+1,\delta s}, \dots, r_{K,\delta s}\}$. Additionally, using the Bayes formula, $p_{k,s} = p(l_{k,\delta s}, r_{\setminus k,\delta s}, Y_{\delta s})$ can be decomposed into the product of $p(Y_{\delta s} | l_{k,\delta s}, r_{\setminus k,\delta s})$, $p(r_{\setminus k,\delta s} | l_{k,\delta s})$ and $p(l_{k,\delta s} | l_{k,s})$. The logarithms of these terms are denoted the *local data driven* term $E_{D,s}$, the *local model driven* term $E_{M,s}$, and the *local smoothing* term $E_{S,s}$, respectively.

The logarithm of (5.2) is referred to as the objective function (OF) and is written

$$\text{as } E(l, Y, l^M) = E_D(l, Y) + E_M(l, l^M) + E_S(l) \quad \text{where} \quad E_D = \sum_{k=1,\dots,K} \sum_{s \in l_k} E_{D,s},$$

$$E_M = \sum_{k=1,\dots,K} \sum_{s \in l_k} E_{M,s} \text{ and } E_S = \sum_{k=1,\dots,K} \sum_{s \in l_k} E_{S,s} \text{ denote the global } \textit{data driven}, \textit{model driven}$$

and smoothing terms, respectively.

Let $A(r_{\setminus k, \delta s}) = a = \{a_1, \dots, a_m\}$ be the set of local labels defined on $r_{\setminus k, \delta s}$ where $A(r_{\setminus k, \delta s})$ denotes the operation of label extraction from the region process $r_{\setminus k, \delta s}$. Regrouping together the sites having similar labels on $r_{\setminus k, \delta s}$ yields the partition $r_{\setminus k, \delta s} = \{r_{\setminus k, \delta s}^{(a_1)}, \dots, r_{\setminus k, \delta s}^{(a_m)}\}$. Let $\bar{Y}_{\delta s_{in(k)}^{(a_j)}}$ and $\bar{Y}_{\delta s_{out(k)}^{(a_j)}}$ denote the mean values of the image Y computed over the l_k defined inner and outer parts, $\delta s_{in(k)}^{(a_j)}$ and $\delta s_{out(k)}^{(a_j)}$, of the a_j labeled part of δs , respectively. The final proposed forms of $E_{D,s}$, $E_{M,s}$ and $E_{S,s}$ are (Mallouche *et al.* 1997a):

$$E_{D,s} = \sum_{j=1, \dots, m} \rho_{s,j} \cdot f \left(\bar{Y}_{\delta s_{in(k)}^{(a_j)}} ; \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right) \quad (5.3)$$

$$E_{M,s} = \sum_{j=1, \dots, m} \rho_{s,j} \cdot g(a_j ; l_{k, \delta s}) \quad (5.4)$$

$$E_{S,s} = \gamma_1 \|l_{k, \delta s}\| + \gamma_2 \|d\theta_{k,s}\|^2 + \gamma_3 \|d^2\theta_{k,s}\|^2 \quad (5.5)$$

For the complete development with discussions and the choice of the functionals f and g , and the coefficient $\rho_{s,j}$, we refer the reader to (Mallouche *et al.* 1997a). Between different

forms, $\rho_{s,j} = \frac{|r_{\setminus k, \delta s}^{(a_j)}|}{|\delta s|}$ presents a convenient tradeoff between the precision of detection and

a local weak repulsive behavior between uncrossing curves. $\rho_{s,j}$ is additionally discussed in Section 5.6.3.

The functional f is given as a *relative* inner-to-outer intensity difference, $f\left(\bar{Y}_{\delta s_{in}(k)}^{(a_j)}; \bar{Y}_{\delta s_{out}(k)}^{(a_j)}\right) = \bar{Y}_{\delta s_{out}(k)}^{(a_j)} - \bar{Y}_{\delta s_{in}(k)}^{(a_j)}$. Here, we assume that the inner region intensity is ordinarily greater than the outer one. The *model driven* functional g is discussed in Section 5.6.4.

5.6.2 Objective function with extremal and inner curves

In the proposed probabilistic criterion (Mallouche *et al.* 1997a) we assumed that all SCCs are extremal. In this work, we assume that each object is made up of subobjects gathered with respect to an intrinsic relational 3-D graph. Additionally, we assume that the semitransparent projection of a subobject generates one extremal SCC and zero, one or two inner curves according to the planar, conic or cylindric nature of the subobject.

Let l_k be an extremal curve and l_{k1} be an inner curve of l_k . l_k and l_{k1} can be partitioned into $l_k = \{l_k^{k1}, l_k^{\setminus k1}\}$ and $l_{k1} = \{l_{k1}^k, l_{k1}^{\setminus k}\}$ where $l_k^{k1} \equiv l_{k1}^k$ is the *double* segment, and $l_k^{\setminus k1}$ and $l_{k1}^{\setminus k}$ are the *simple* segments. Any distribution $p(l_k, l_{k1}, \cdot)$ can be rewritten as $p(l_k^{k1}, l_k^{\setminus k1}, l_{k1}^k, l_{k1}^{\setminus k}, \cdot) = p(l_k^{k1}, l_k^{\setminus k1}, l_{k1}^{\setminus k}, \cdot) = p(l_k^{\setminus k1}, l_{k1}^k, l_{k1}^{\setminus k}, \cdot)$ since both segments l_k^{k1} and l_{k1}^k represent the same physical event. Consequently, we redefine the sum $\sum_{s \in l_k}$ in E_D , E_M

and E_S as $\sum_{s \in S(l_k)}$ where $S(l_k) = l_k$ when l_k is an extremal curve and $S(l_{k1}) = l_{k1}^{\setminus k}$ when l_{k1} is

an inner curve of l_k . This adjustment is made in order to prevent biasing the solution toward the *double* curves and to preserve the precision of the detection.

5.6.3 Data driven term

The *data driven* term is given by:

$$E_D(l, Y) = \sum_{k=1, \dots, K} \sum_{s \in S(l_k)} \sum_{j=1, \dots, m} \rho_{s,j} \cdot \left(\bar{Y}_{\delta s_{in(k)}^{(a_j)}} - \bar{Y}_{\delta s_{out(k)}^{(a_j)}} \right) \quad (5.6)$$

Let us assume that $\rho_{s,j}$ is fractional defined by $\rho_{s,j} = \frac{\mu_{s,j}}{v_s}$. For the simple model with

extremal curves we choose $\rho_{s,j} = \frac{|r_{\setminus k, \delta s}^{(a_j)}|}{|\delta s|}$. More precisely, the numerator is given by

$$\mu_{s,j} = |r_{\setminus k, \delta s}^{(a_j)}| = |\delta s_{in(k)}^{(a_j)}| + |\delta s_{out(k)}^{(a_j)}| \text{ if both sets } \delta s_{in(k)}^{(a_j)} \text{ and } \delta s_{out(k)}^{(a_j)} \text{ are not empty. Otherwise, } \mu_{s,j}$$

is set to zero. We now examine the expression of $\rho_{s,j}$ when the model contains inner curves

besides extremal curves. We refer the reader to Fig. 5.3 that shows three sets of local views of an extremal-inner curve junction. The patterns a0, b0 and c0 of the first row show the neighborhood partitioning of δs used for the computation of the local criterion when s

belongs to the *simple* extremal segment l_k^{k1} of l_k , the *double* segment l_k^{k1} of l_k , and the *simple* segment l_{k1}^{k1} of the inner curve l_{k1} , respectively. The second row (a1, b1 and c1) and the third row (a2, b2 and c2) show the same curve pattern with one additional curve l_h crossing l_{k1} and approaching l_{k1} , respectively. We note that when $s \in l_k$, $s \in l_{k1}$ or $s \in l_k^{k1}$, the label partition $\alpha = A(r_{\cdot, \delta_s})$ is inferred from $r_{\setminus k, \delta_s}$, $r_{\setminus k1, \delta_s}$ or $r_{\setminus k \setminus k1, \delta_s}$, respectively. When generating the region process $r_{\setminus k \setminus k1, \delta_s}$ for the *double* segment l_k^{k1} , segments l_k and l_{k1} are considered as a single curve. This implies that after extracting the set of labels $\alpha = A(r_{\setminus k \setminus k1, \delta_s})$ each $r_{\setminus k \setminus k1, \delta_s}^{(a_j)}$, $j=1, \dots, m$, is partitioned into two parts: (i) the outer part $\delta s_{out(k)}^{(a_j)}$ using l_{k, δ_s} , and the inner part $\delta s_{in(k1)}^{(a_j)}$ using l_{k1, δ_s} . The second column of Fig. 5.3 illustrates these operations.

Let $\mu_s = \sum_{j=1, \dots, m} \mu_{s,j}$ be the sum of the weighting coefficients $\mu_{s,j}$ used while computing

the *local data driven* term $E_{D,s}$ for the site $s \in l_k$. For $\rho_{s,j} = \frac{|r_{\setminus k, \delta_s}^{(a_j)}|}{|\delta s|}$, the difference between

the *complete* model with (MW) inner curves and the *simple* model without (MWO) inner curves is that for MWO, μ_s is almost equal to $|\delta s|$ (Mallouche *et al.* 1997a) whereas μ_s can significantly decrease for MW in the regions where extremal and inner curves exist together.

In Fig. 5.3, μ_s is given by the sum of the shadowed regions in each pattern of the first row and second row. Consequently, for such patterns, $E_{D,s}$ can decrease due to the reduction in the total μ_s . This variation in $E_{D,s}$ relatively to different patterns can be considered as an important drawback. We can remedy to this by replacing the constant denominator v_s in $\rho_{s,j}$ by the following:

$$v_s = \begin{cases} |\delta s \cap out(l_{k1})|, & \text{if } s \in l_k \text{ extremal} \\ |\delta s \cap [out(l_k) \cup in(l_{k1})]|, & \text{if } s \in l_k \text{ extremal and } s \in l_{k1} \text{ inner} \\ |\delta s \cap in(l_k)|, & \text{if } s \in l_{k1} \text{ inner} \end{cases} \quad (5.7)$$

Columns from left to right in Fig. 5.3 show three different patterns for $s \in l_k$, $s \in l_k^{k1}$ and $s \in l_{k1}$, respectively. The sum of surfaces of all shaded regions in each pattern gives the

corresponding v_s . Note that when v_s is given by (5.7) $\rho_{s,j} = \frac{\mu_{s,j}}{v_s}$ provides the same

sensitivity, precision, coherence, robustness and weak repulsion properties for the *data driven* term in both MW and MWO.

5.6.4 Model driven term

In this section, we discuss and develop the *model driven* functional $g(a_j; l_{k,\delta s})$. g is the measurement of the *a priori* conditional distribution $p(r_{\setminus k, \delta s} | l_{k, \delta s})$. This distribution can be

considered as a *local-likelihood* measurement or a local self consistency measurement of the solution. More precisely, $p(r_{\setminus k, \delta_s} | l_{k, \delta_s})$ measures the consistency with respect to the model M between l_{k, δ_s} and the rest of the curves $r_{\setminus k, \delta_s}$. The *a priori* model M can include the geometric or relational representation, or any statistical knowledge about the object to recognize. We recall the *global model driven*:

$$E_M(l, M) = \sum_{k=1, \dots, K} \sum_{s \in S(l_k)} \sum_{j=1, \dots, m} \rho_j \cdot g(a_j; l_{k, \delta_s}; M) \quad (5.8)$$

The presence of E_M in the objective function introduces a *top-down* influence. The high-level *a priori* knowledge constraints the search process and hence reduces the dimension of the search space.

5.6.4.1 What kind of solution-model consistency measurement?

We now address the question of the specification of the local functional $g(a_j; l_{k, \delta_s}; M)$. g measures the consistency of l_{k, δ_s} when $l_{k, s}$ belongs to the curve segment $l_k^{(b_i)} \subset l_k$ separating two adjacent regions. Fig. 5.4 illustrates the curve segmentation operation of two superposed free-forms, conic $\{l_0, l_1\}$ and planar l_2 , into eight curve segments. l_{k, δ_s} can be locally described by its *curve label* b_i and its direction $\theta_{k, s}$, curvature $d\theta_{k, s}$ and the higher order derivatives $d^n \theta_{k, s}$, $n=2, 3, \dots$. Let $v_{k, s} = \{b_i, \theta_{k, s}, d\theta_{k, s}, d^2 \theta_{k, s}, \dots\}$ be the parameter vector of the local segment $l_{k, s}$. The position (x_s, y_s) of site s is intentionally discarded from $v_{k, s}$

in order to avoid introducing elastic behavior in g (Terzopoulos *et al.* 1987). Using this representation, g can be considered as a function of $v_{k,s}$ and M and will be denoted by $g(b_i, \theta_{k,s}, d\theta_{k,s}, d^2\theta_{k,s}, \dots; M)$.

Let the model M be given as segmented multiple curve $l^M = \{l_k^{M(b_i)}\}$. g can also be written as $g(v_{k,s}, v_{k,t}^M)$, where $s \in l_k^{(b_i)}$, $t \in l_k^{M(b_i)}$ and sites s and t are related by a one-to-one mapping function between solution l and model M . g can be designed using perceptual organization criteria (Grimson 1990; Horaud and Skordas 1989). A simple expression for g is the distance $\|v_{k,s} - v_{k,t}^M\|$. The drawbacks of this representation are the updating computational requirements of the one-to-one mapping function along the search process and the restrictive one-to-one measurement $g(v_{k,s}, v_{k,t}^M)$ that can heavily constrain the search process and hence reduce its exploration capability. In order to avoid these drawbacks, we propose, in this work, a statistical approach to develop the solution-model consistency measurement.

5.6.4.2 Statistical representation

The basic idea of our technique is to replace the geometric model by its region adjacency (relational) graph augmented by the statistical knowledge of the model. Let $G^M(A^M, B^M)$ be the relational graph defining the region neighboring relation of the 2-D model M .

$A^M = \{a_j, j=1, \dots, J\}$ and $B^M = \{b_i, i=1, \dots, I\}$ denote the set of nodes (vertices) and the set of branches of the graph G^M , respectively. Each node represents some connected region of the 2-D model M , and each branch represents a boundary segment separating two adjacent regions. Thus, the set B^M can be seen as the projection of the segmented curve set $I^M = \{I_k^{M(b_i)}\}$ into the relational graph G^M . The idea now is to augment each branch $b_i \in B^M$ by a weighted sum of the statistical distributions of the local parameters of its curve segment $I_k^{M(b_i)} \in I^M$. This means that the informational geometric content of $I_k^{M(b_i)}$ is compressed in some global representation.

$$g(b_i, \theta, d\theta, d^2\theta, \dots) = \sum_{n=0,1,\dots} c_n \cdot H_n^{G_\sigma}(I_k^{M(b_i)}, d^n\theta) \quad (5.9)$$

where $H_n(I_k^{M(b_i)}, d^n\theta)$ denotes the histogram distribution of the parameter $d^n\theta$ of curve segment $I_k^{M(b_i)}$ and $c_n, n=0, 1, \dots$, denote the weighting coefficients. The Gaussian filtered histograms $H_n^{G_\sigma}(I_k^{M(b_i)}, \cdot) = G_\sigma(\cdot) * H_n(I_k^{M(b_i)}, \cdot)$, $i=1, \dots, I$ and $n=0, 1, \dots$ can be used in order to reduce the fluctuation effect caused by the discretization, and the standard deviation σ can be experimentally defined.

However, this may produce a modulation (bias) of each solution segment $I_k^{(b_i)}$ by the statistical knowledge $\{H_n^{G_\sigma}(I_k^{M(b_i)}, \cdot)\}$ that can heavily compromise the precision of the

discontinuity detection specially for regions with low S/N. In order to prevent this effect, we replace histograms $H_n^{G_o}(I_k^{M(b_i)}, \cdot)$ by trapezoidal thresholded functions $WH_n^{G_o}(I_k^{M(b_i)}, \cdot)$ (or pseudo-binary):

$$WH_n^{G_o}(I_k^{M(b_i)}, d^n\theta) = \min_{-\Delta_n \leq \delta \leq \Delta_n} \left\{ \frac{|\delta| - \Delta_n}{\Delta_n} \cdot \text{true} \left(H_n^{G_o}(I_k^{M(b_i)}, d^n\theta + \delta) > \text{Th}_n \right) \right\} \quad (5.10)$$

δ denotes the scanning variable bounded to $[-\Delta_n, \Delta_n]$ and belonging to the abscissa variable $d^n\theta$ of $H_n^{G_o}$. The binary function true returns one if its argument is true, otherwise it returns zero.

For each histogram $H_n^{G_o}(I_k^{M(b_i)}, \cdot)$, the threshold Th_n and the linear trapezoidal side Δ_n of the generated $WH_n^{G_o}(I_k^{M(b_i)}, \cdot)$ are parts of the model. The idea here is to introduce a fixed high-level influence through the *model driven* term when the local value of $d^n\theta$ is within the *a priori* window opening of $WH_n^{G_o}(I_k^{M(b_i)}, \cdot)$. This influence linearly vanishes when the parameter value goes away from the window opening. The choice of Th_n and Δ_n must account for the discretization effect, the deviation and the variation of the model.

In addition, we extend the scanning length of each $I_k^{M(b_i)}$ by some relative length λ_k in both *CW* and *CCW* directions. So, we replace $I_k^{M(b_i)}$ in $WH_n^{G_o}(I_k^{M(b_i)}, \cdot)$ by $I_k^{M(b_i)} \upharpoonright \lambda_k$ denoting the extended segment. This is helpful in some specific situations where, for example,

$WH_0^{G_\sigma}(I_k^{M(b_i)}, \theta)$ is uninodal while $WH_0^{G_\sigma}(I_k^{M(b_i)} \upharpoonright \lambda_k, \theta)$ is multinodal due to an abrupt change in the curve characteristics inside the extension parts $\upharpoonright \lambda_k$. In this case, the *model driven* term is more constraining with $WH_0^{G_\sigma}(I_k^{M(b_i)}, \theta)$ than with $WH_0^{G_\sigma}(I_k^{M(b_i)} \upharpoonright \lambda_k, \theta)$ and hence can create a *gluing* effect that in certain situations obstructs convergence toward the solution. Fig. 5.5 illustrates a typical situation. λ_k can depend on the relative deformation of the solution or the mutual curve displacement from the model. Practically, we choose $\lambda_k = \frac{\sigma^M}{2\pi} \|I_k^M\|$ where σ^M quantifies the variance of the object population represented by the geometric model I^M .

5.7 Search method

Starting from an initial wire-frame estimate (multiple SCCs), the search method iteratively deforms the solution in order to optimize the objective function. The deformation process is also referred to as the matching process. We assume that the initial state is consistent with the selected model and hence has the same 2-D relational graph. Search methods by optimization of a single global criterion may be either deterministic or stochastic. The choice of the optimization method is directly related to the convexity level of the search space that is by its turn related to the transition process between states. A rich transition process engenders a high connectivity and consequently a dense neighborhood in the state space. This can reduce

the number of local optima and consequently enhance the convexity level of the search space.

Many techniques can be used to create a richer transition process. Long distance deformations during the search process produce transitions between geometrically distant states. In addition, simultaneous curve deformations generate a richer transition process. The transition process can be additionally enhanced by introducing long distance interactions such as, e.g., the multiresolution technique that concentrates both the observed image and the model in a pyramidal scheme (Terzopoulos 1986).

Because our OF is not convex and we want to avoid computationally burdensome stochastic methods, we choose to develop an affordable descent search method with randomized sampling capable to overcome many local optima. The randomized sampling allows us to create a rich system of state transition and thus to improve the convexity of the search space. Our approach consists of gradually deforming the initial state until reaching acceptable solution, and of further reducing the non convexity of the search space by using long range deformations. The structure of our descent method is presented in Fig. 5.6. The method is made up of three embedded levels: (i) the global affine alignment, (ii) the multiresolution structure, and (iii) the progressive deformation process.

First, we perform global alignment of the initial solution to the observed image. This allows us to adjust the initial solution as close as possible to the object in the image. Our technique of global alignment is iterative and based on a dynamic model that manipulates the solution using an affine transformation in order to optimize a similarity objective function (Mallouche *et al.* 1997b).

The remaining part of the search process is embedded in a multiresolution technique. This

technique consists of building a pyramidal structure organized from fine-to-coarse resolution (Terzopoulos 1986). First, we solve for the coarsest resolution. The resulting solution is projected onto the next finer level, and the search is repeated. This process is iterated until reaching the finest resolution. The multiresolution technique creates long range interactions and accelerates the convergence.

The progressive deformation process is centered on the new state generation (NSG) or the state sampler that is the extremal curve and/or inner curve deformation unit. This unit generates a new state by randomly deforming the current state. Ideally, a sample should be generated from simultaneous multiple deformations. These can be randomly applied to some randomly selected curves. The strategy can produce a very rich transition process. However, it is computationally burdensome when implemented on a sequential calculator. In this instance, we choose a simple deformation technique that consists of randomly deforming one curve or possibly a pair of extremal-inner curves at each iteration. The maximum distance of deformation is limited by a relaxed maximum value.

The next step is the computation of the criterion value variation (CVV). It consists of scanning all the affected sites by the last deformation and then computing the summation of the local criterion value. The difference between the new value and the old value of the criterion summation of affected sites is called the CVV. The new generated state is accepted if the CVV is negative. Otherwise, the previous state is kept. The acceptance decision could be made stochastic by allowing transition to less probable state.

We now detail the deformation technique, the algorithmic constraints for model preservation and the functional constraint.

5.7.1 Deformation technique

Our search method is based on matching a free form deformable initial state with the observed image. The matching is a progressive process of randomized deformations that are continuously controlled by algorithmic constraints and functional constraint. The functional constraint corresponds to the acceptance criterion of the chosen method to optimize the objective function. The algorithmic constraints ensure the preservation of the multiple SCCs representation of the solution and the topological structure (relational content) of the solution that are denoted the SCC preservation (SCCP) and relational model preservation (RMP), respectively. SCCP can be further detailed by three conditions: (i) The deformation has to preserve connectivity, and closure; (ii) The deformation must not generate a curve crossover; (iii) The deformation has to preserve curve orientation; We show that RMP can be translated into two conditions: (i) The old segment and the new segment have to have same descriptions; (ii) The region filling and erasing generated by the curve deformation have to be consistent with the model.

The deformation technique is formed of two sequential processes (Fig. 5.6): (i) Sampling and curve segment generation; (ii) Region updating (filling and erasing) that directly translates the line-region duality (Section 5.5.1). The first process integrates the SCCP conditions and the first RMP condition, while the second process integrates the second RMP condition.

The sampling and curve generation process consists of (i) selecting the curve and then delineating the segment to be deformed, and (ii) generating the new curve segment to replace the old one.

5.7.1.1 Segment to deform

The first step is achieved using two sampling processes and a bidirectional curve scanning. The first sampling process consists of randomly selecting a site s of L^I . All curves that pass through s will be sequentially deformed. The site s is used as a reference point for sampling the summit point of deformation, denoted s_a . Then, it is used to scan and find the curve segment to be deformed $l_{k,(s_0,s_1)}$ where s_0 and s_1 are the extremities of $l_{k,(s_0,s_1)}$ in the *CW* direction. The point s_a is sampled inside a disk $\Delta_{s,d}$ centered on site s and with radius d , called the maximum deformation distance (Fig. 5.7) according to a pre-specified distribution. We note that the search radius d can be decreased to allow a progressive reduction of the range of deformation in order to refine the solution. We now delineate the curve segment $l_{k,\Delta_{s_a,d}} = l_k \cap \Delta_{s_a,d}$. The curve segment $l_{k,(s_0,s_1)}$ is then defined as the largest connected segment of $l_{k,\Delta_{s_a,d}}$ offering the largest viewing angle from s_a . Fig. 5.8 shows two typical detections of $l_{k,(s_0,s_1)}$ when s_a belongs to the outer or the inner region of l_k , respectively.

This detection procedure is applicable for both extremal and inner curves. However, if the curve to deform l_k is extremal and $l_{k,(s_0,s_1)}$ has a common portion with one of its inner curve l_h , we consider deforming both $l_{k,(s_0,s_1)}$ and l_h . Consequently, we extract the segment to be deformed $l_{h,(t_0,t_1)}$ from l_h and the joint point (*splitting* or *merging*) of l_k with l_h denoted

s_{SM} . Fig. 5.9 shows a *splitting* pattern where $l_{k,(s_0,s_{SM})}$ and $l_{h,(t_0,s_{SM})}$ are *double* segments representing the same physical reality.

5.7.1.2 Generation of the new segment

Two patterns have to be considered in generating the new curve segment: (i) deforming a *simple* curve, and (ii) deforming connected curves. We denote l_k the sampled curve to be deformed and l_h the possibly connected curve to l_k . Here, we assume that if l_k is inner curve, then it can be deformed alone as a simple extremal segment.

For the *simple* curve deformation, the new segment $l_{k,(s_0,s_a,s_1)}$ replacing $l_{k,(s_0,s_1)}$ is generated using two glued pieces of cubic splines denoted $l_{k,(s_0,s_a)}$ and $l_{k,(s_a,s_1)}$ (Fig. 5.10).

This scheme preserves the zero and first order continuities of the deformed curve.

For the deformation of connected curves (l_k, l_h) where l_k is the extremal curve and l_h is the inner curve three schemes are considered according to the existence of zero, one or two junction points between the connected segments. Fig. 5.11 illustrates the deformation of a pair of connected curves, extremal $l_{k,(s_0,s_1)}$, and inner $l_{h,(t_0,t_1)}$, having one *splitting point* t_{SM} . In order to preserve the zero and first order continuities and the inner-outer topologic relation and the geometric aspect of the local connected curves (l_k, l_h) , (l_k, l_h) are deformed together as follows:

- The extremal segment is deformed exactly as for *simple* curve. The new extremal segment

is $l_{k,(s_0,s_a,s_1)}$.

- The position of the split point t_{SM} on the old segment $l_{k,(s_0,s_1)}$ is used to compute the

$$\text{ratio } \tau = \frac{\|l_{k,(s_0,s_{SM})}\|}{\|l_{k,(s_0,s_1)}\|}.$$

- The position of the new split point s_{aSM} on $l_{k,(s_0,s_a,s_1)}$ is given by

$$\|l_{k,(s_0,s_{aSM})}\| = \tau \|l_{k,(s_0,s_a,s_1)}\|.$$

- The new segment $l_{h,(t_0,t_{aSM},t_1)}$ of the inner curve l_h is formed of the two glued segments $l_{h,(t_0,t_{aSM})}$

and $l_{h,(t_{aSM},t_1)}$ where $l_{h,(t_0,t_{aSM})}$ is exactly similar to $l_{k,(s_0,s_{aSM})}$ of $l_{k,(s_0,s_a,s_1)}$ and $l_{h,(t_{aSM},t_1)}$

is a cubic spline. Meanwhile, in order to preserve the topologic definition of the inner

curve, $l_{h,(t_{aSM},t_1)}$ is constrained to belong to $\bar{in}(l_k) = in(l_k) \cup l_k$. We here assume that the

region process of l_k (filling-erasing) has already been updated.

Equivalently, for one *merging* point of connected curves, the curve generation uses the same algorithm as for a single *split* point with one exception. The generated inner segment is given as $l_{h,(t_0,t_{aSM},t_1)} = l_{h,(t_0,t_{aSM})} \cup l_{h,(t_{aSM},t_1)}$ where $l_{h,(t_0,t_{aSM})}$ is the cubic spline and

$$l_{h,(t_{aSM},t_1)} = l_{k,(s_{aSM},s_1)}.$$

For connected segments with two joint points, we apply the same procedure as for the

single joint point. For connected segments without joint points, first, we generate the extremal segment from which the inner segment is directly copied.

At this level, the generated new segment has to replace the old segment. Still, before doing this, the old segment has to be buffered in order to restore the initial state if the generated deformation is eventually rejected. However, for connected segments, inner curve handling is delayed until checking that the new extremal segment fulfills the algorithmic constraints and updating the corresponding region process.

Finally, the preservation of the *continuity* (closure) and *orientation* of SCC is assured by using a sequentially coded chain for the SCC discrete representation, where each sequentially generated site is the neighbor of the previously generated one. The *uncrossing* condition for SCC preservation can also be integrated at this level using a *look ahead* detection of the occupied sites while sequentially generating the new segment.

5.7.2 Algorithmic constraints

Two important aspects of the proposed search technique are the SCCP and the RMP. On one hand, the SCCP is intrinsically integrated in the curve generation sampler as previously detailed. On the other hand, the RMP has to ensure that the new generated sample (NGS) does not modify the relational aspect (or the topologic content) of the initial solution. This verification is referred to as the *algorithmic constraints*. Equivalently, each NGS that does not meet the algorithmic constraints is rejected. For this purpose, both of the NGS δl and its corresponding variation δr of the region process have to fulfill preservation conditions.

We now detail these conditions.

Note that model l^M and initial state l^0 have the same 2-D relational graph $G^M(A^M, B^M)$. The search process accepts any initial state if its 2-D relational graph $G(A, B)$ is identical to G^M . In this sense, we say that two different states are relationally identical. Recall that each vertex $a_j \in A$ of $G(A, B)$ represents one connected region labeled by a_j and each branch $b_i \in B$ represents the boundary segment between two adjacent regions.

Let A be the operation of region label extraction. When applied to line processes, A provides a sequential list of labels, while when applied to region processes, it provides a set of labels

For a region partitioned plane, the region adjacency graph and its dual graph are *planar* which means that they lie in a plane without intersecting branches (Ballard and Brown 1982). Hence, the 2-D region relational graph $G(A, B)$ of any multiple *simple* curve state l is *planar*.

We now state the proposition of the relational graph preservation.

Proposition 1: *A curve deformation $\delta l_k = l_{k(s_0, s_1)} - l_{k(s_0, \dots, s_1)}$ of a multiple simple closed curve state $l = \{l_k, k = 1, \dots, K\}$ preserves the 2-D relational graph $G(A, B)$ of l if it fulfills two conditions:*

1- *The sequential description $A_{\setminus k}(l_{k(s_0, \dots, s_1)})$ over $r_{\setminus k}$ exactly matches $A_{\setminus k}(l_{k(s_0, s_1)})$;*

2- *The set of labels $A_{\setminus k}(\delta r_k)$ over $r_{\setminus k}$ of the region variation δr_k is a subset of the set*

$$A_{\setminus k}(l_{k(s_0, s_1)})$$

Proof: Let $G_{\setminus k}$ be the relational graph of the region process $r_{\setminus k}$. Note that deforming l_k does not affect $G_{\setminus k}$. The superposition of l_k on $r_{\setminus k}$ transforms $G_{\setminus k}$ into G . We want to prove that if the new segment $l_{k,(s_0 \dots s_1)}$ crosses $r_{\setminus k}$ in the same sequential manner as $l_{k,(s_0, s_1)}$ and fulfills condition 2, δl_k preserves G . Let $A_{\setminus k}(l_{k,(s_0, s_1)}) = \{a_{k,1}, a_{k,2}, \dots, a_{k,n}\}$ be the sequential label list (curvilinear scanning from s_0 to s_1) of the regions of $r_{\setminus k}$ crossed by $l_{k,(s_0, s_1)}$ where $a_{k,i}$, $i=1, \dots, n$, is a vertex of $G_{\setminus k}$. $a \in G_{\setminus k}$ represents the label of the connected region $r_{\setminus k}^{(a)}$ of $r_{\setminus k}$. Any $r_{\setminus k}^{(a)}$ can either be a simply or a multiply connected region. In the latter case, this is equivalent to have one (or more than one) subgraph $sG_{\setminus k} \subset G_{\setminus k}$ linked to the remaining part of $G_{\setminus k}$ only through vertex a . $sG_{\setminus k}$ is then denoted *isolated* subgraph, and the corresponding region can be seen as a hole in the region $r_{\setminus k}^{(a)}$. It is easy to see that if both $l_{k,(s_0, s_1)}$ and $l_{k,(s_0 \dots s_1)}$ cross only simply connected regions and condition 1 is met, then $l_{k,(s_0, s_1)}$ and $l_{k,(s_0 \dots s_1)}$ follow the same path over $r_{\setminus k}$ and hence δl_k does not modify G (Fig. 5.12, left scheme). If both $l_{k,(s_0, s_1)}$ and $l_{k,(s_0 \dots s_1)}$ cross one (or more) multiply connected region $r_{\setminus k}^{(a)}$ with one (or more) subgraph $sG_{\setminus k} \subset G_{\setminus k}$, two situations may occur: (i) if both $l_{k,(s_0, s_1)}$ and $l_{k,(s_0 \dots s_1)}$ do not cross any region corresponding to $sG_{\setminus k}$ and condition 1 is verified then condition 2

ensures that $l_{k,(s_0,\dots,s_1)}$ has the hole corresponding to $sG_{\setminus k}$ on the same side as for $l_{k,(s_0,s_1)}$ (Fig. 5.12, center scheme); (ii) if both $l_{k,(s_0,s_1)}$ and $l_{k,(s_0,\dots,s_1)}$ cross one (or more) region of the hole corresponding to $sG_{\setminus k}$ and condition 1 is verified then condition 2 ensures that $l_{k,(s_0,\dots,s_1)}$ has the non-crossed regions of the hole corresponding to $sG_{\setminus k}$ on the same side as for $l_{k,(s_0,s_1)}$ (Fig. 5.12, right scheme). Hence, any deformation δl_k verifying conditions 1 and 2 preserves the relational graph $G(A, B)$ of l . ■

Proposition 1 can be also proved for deformations including double segments. In this case, we partition the sequential description $A_{\setminus k}(l_{k,(s_0,s_1)})$ into sequential sublists corresponding to the *simple* and *double* segments of $l_{k,(s_0,s_1)}$. This means that in addition to the label description in $r_{\setminus k}$, the operation $A_{\setminus k}$ eventually extracts the label of the curve (extremal or inner) connected to l_k .

Conditions 1 and 2 are denoted the algorithmic constraints for relational graph preservation. First, if condition 1 is satisfied after generating the deformation δl_k , we launch a filling-erasing procedure (FEP) to generate δr_k and to update region processes r_k and r . Then, condition 2 is tested: if $A_{\setminus k}(\delta r_k) \neq A_{\setminus k}(l_{k,(s_0,s_1)})$ the NGS is rejected and the previous state is kept. Otherwise, the NGS fulfills the algorithmic constraints and has to be tested for the functional constraint.

The technique of the FEP makes direct use of the line-region duality to infer δr_k from δl_k (Section 5.5.1). The FEP is also used to detect the set of curve segments $l \cap \delta r_k$ affected by the region deformation δr_k . $l \cap \delta r_k$ is used to temporarily update the curve labels of all sites belonging to $l \cap \delta r_k$, and then to compute the criterion value variation caused by the deformation δl_k .

5.7.3 Functional constraint - Criterion value variation

We now have to compute the CVV or the variation δE of the objective function E caused by the NGS δl_k . Here, δl_k verifies the algorithmic constraints. E is spatially distributive, thus δE equals to the variation of E over the affected sites. We have three sets of affected sites. The first set and the second set consist of sites directly affected by the deformation, and are given by $AS1(\delta l_k) = l_{k,(s_0 \dots s_1)} \cup l_{k,(s_0, s_1)}$ and $AS2(\delta l_k) = l \cap \delta r_k$, respectively. The local criterion value E_s is computed for each $s \in l_{k,(s_0 \dots s_1)}$, as shown in Section 5.6. The CVV due

to $AS1(\delta l_k)$ is simply given as $\delta E_{AS1} = \sum_{s \in l_{k,(s_0 \dots s_1)}} E_s^N - \sum_{s \in l_{k,(s_0, s_1)}} E_s$, where E_s^N denotes the new

local energy computed after δl_k . The CVV due to $AS2(\delta l_k)$ is given as

$\delta E_{AS2} = \sum_{s \in l \cap \delta r_k} E_s^N - E_s$. The third set $AS3(\delta l_k)$ of indirectly affected sites is given by

$AS3(\delta I_k) = \{s \mid \exists t \in \delta s; t \in \delta r_k\}$. $AS3(\delta I_k)$ can be generated while computing δE_{AS1} and δE_{AS2} as follows:

$$AS3(\delta I_k) = \{t \in \delta s \text{ and } t \in AS1(\delta I_k) \cup AS2(\delta I_k) \mid s \in AS1(\delta I_k) \cup AS2(\delta I_k)\} \quad (5.11)$$

Finally, we compute $\delta E_{AS3} = \sum_{s \in AS3(\delta I_k)} E_s^N - E_s$.

The last step is the acceptance decision. The CVV $\delta E = \delta E_{AS1} + \delta E_{AS2} + \delta E_{AS3}$ is used to decide whether to accept or to reject the NGS. The decision can be either stochastic or deterministic as indicated earlier in this paper. In this work, we use a deterministic decision.

5.8 Implementation and experimental results

5.8.1 Implementation

The implementation was written in C language for sequential processors. However, the approach can also be implemented on parallel machines because of the distributive form of the global criterion. For all experiments, the *a priori* model and the initial solution, I^M and I^0 were chosen to be identical. The implementation is made up of three parts: (i) The initialization to create the data base with the line-region duality and to infer the model relational graph and its corresponding attributes; (ii) The global alignment of the initial solution; (iii) Finally, the matching process with the iterative deformation technique.

In the first part, I^M and I^0 are formed of the superposition of K SCCs. Both region and

line processes are defined over homogeneous square grid lattices. Each SCC is discretized as a closed bidirectional coded chain. Both extremal and inner curves have the same data structure.

Then, the relational graphs $G^M(A^M, B^M)$ of I^M and $G^0(A^0, B^0)$ of I^0 are inferred and checked to be identical. A planar relational graph G can be rewritten as a sequential series of curve segments or branches where each branch represents a boundary segment separating two adjacent regions. For this, the K curves are orderly scanned and then each I_k is curvilinearly scanned in a CW direction. The segment label is incremented each time the local descriptors of current site s_t and previous site s_{t-1} are different. Finally, the histograms $WH_0^{G^0}(I_k^{M(b,i)}, \theta)$ (*model driven* functions) are computed by scanning again the curve segments $I_k^{M(b,i)}$ of I^M extended by the additional curve portions λ_k . In the current implementation, we only use a zero order angular parameter θ for the *model driven* term.

The second part is the global alignment. It adjusts the initial solution as close as possible to the real object. This step is important when using descent search method and when the initial state is far from the real solution. The basic idea of this technique detailed in (Mallouche *et al.* 1997b), is to manipulate an affine transformation applied to the initial state in order to optimize some similarity measurement between the transformed state and the observed image. The aligned state then becomes the initial state of the deformation process.

The third part is the matching process. It randomly deforms the solution to optimize the

objective function $E(I, Y, I^M)$ (Section 5.7). Fig. 5.6 shows the structure of the search method. Each iteration consists of the following five steps: (i) Randomly select a curve l_k and then a center site s_a to where the curve will be deformed. (ii) Determine the segment to be deformed $l_{k,(s_0,s_1)}$ and its corresponding descriptor. (iii) Generate the new segment $l_{k,(s_0,s_a,s_1)}$ and its descriptor, then check the first algorithmic constraint. (iv) Apply the FEP and verify the second algorithmic constraint. (v) Finally, compute the value variation δE of the objective function to decide accepting or rejecting the proposed deformation. These steps are repeated until reaching a prefixed iteration number (terminating condition). We note that the matching process is embedded in a multiresolution representation (Section 5.7).

5.8.2 Results

Our multiple curve matching technique was tested with noisy synthetic images and noisy real images of a synthetic object. Two sets of synthetic images were generated to simulate X-ray projections of free form conic and cylindric objects. Two conic and cylindric hand-drawn free form models (Fig. 5.13) were used to match the generated images. Images were generated to simulate objects having higher intensity boundaries as, e.g., X-ray images of skeletal bones. The boundary and center intensities were set to 1.0 and 0.2, respectively. The 2-D semitransparent projections of the free form conic and cylindric objects are shown in panel $a0$ of Fig. 5.14 and Fig. 5.15, respectively. Images $a1$ and $a2$ were generated from $a0$ by adding an i.i.d. noise $N(\mu=0; \sigma)$ of $\sigma=0.5$ and 1.0, respectively. Panels $b0$, $b1$ and $b2$ of Fig.

5.14 and Fig. 5.15 show the results of matching the corresponding models with images α_0 , α_1 and α_2 of both figures, respectively. Results are shown as region and curve coded images. Experiments proved robustness with respect to noise of our approach. Intensity discontinuities were precisely detected in spite of the weak S/N. Note that the models (Fig. 5.13) (and consequently the initial states) were far away and different from the real objects. We also note the coherent coexistence between extremal and inner curves.

Results of matching the hand-drawn models a and b of Fig. 5.16 with real images are presented in Figures 17, 18 and 19. Models a and b of Fig. 5.16 have different relational graphs. Images α_0 , α_1 and α_2 of Fig. 5.17 and Fig. 18 are two sets of X-ray projections of the same 3-D synthetic object (SO). The SO is an acrylic handmade object having the overall dimensions of $(45 \times 50 \times 75)$ mm. SO was particularly designed for modeling a human vertebra. Five subobjects were used to assemble the SO: (i) a free cylindric form for the vertebral body; (ii) two smaller cylindric forms for the pedicles; (iii) a butterfly planar form for the apophysis; and finally, (iv) a small free conic form for the spine. The acrylic structure was then coated with 2 to 5 mm of plaster mixed with 10 percent of Barium. X-ray images were produced under 85kV and 32mAs setting while the SO was immersed in a water filled cylindric container. X-ray images were finally digitalized with 2^{10} gray levels and then windowed in 128×128 pixels. The first row in each of Fig. 5.17 and Fig. 5.18 is composed of three X-ray images, α_0 , α_1 and α_2 , produced from three different viewing positions while approximately preserving the same 2-D relational graph. These images are of high complexity with low and changeable contrast basically due to the complex 3-D composition of the phantom and to the semitransparent property of X-ray penetration. The model a of Fig. 5.16

was used to match with a_0 , a_1 and a_2 of Fig. 5.17 where b_0 , b_1 and b_2 are the corresponding results. The model b of Fig. 5.16 was used to match with images a_0 , a_1 and a_2 of Fig. 5.18 where the corresponding results are b_0 , b_1 and b_2 , respectively. Both matching series produced excellent results. We note the coherent convergence between the different SCCs (extremal and inner) constituting the solution in different crossing and non-crossing situations. The precision of detection is remarkably high even for regions with weak and variable contrast and variable background.

The final set of experiments aims at testing the approach behavior with extremely noisy images. The experiment consisted of matching the model a of Fig. 5.16 with the three noise-corrupted images, a_0 , a_1 and a_2 , of Fig. 5.19. Images were generated from a_0 of Fig. 5.17 by adding an i.i.d. Gaussian noise $N(\mu=0; \sigma)$ with $\sigma=0.1$, 0.2 and 0.3 , respectively. b_0 , b_1 and b_2 of Fig. 5.19 are the corresponding results of a_0 , a_1 and a_3 , respectively. Results confirmed the stability and the robustness of the approach with respect to noise.

5.9 Discussion and conclusions

Our model-based approach for matching with non rigid objects is characterized by the following points: (i) Multiple simple closed curve representation for modeling the semitransparent image formation; (ii) Structural geometric modeling using simple planar, conic and cylindric free form objects; (iii) A topologic framework to integrate extremal and inner curves modeling the semitransparent projection of the free form objets; (iv) A probabilistic framework to develop the objective function; (v) Algorithmic constraints for

preserving the 2-D relational graph of the start solution; (vi) The distributive form of the global criterion and its local dependency to the observed image and the *a priori* model; (vii) Line-region duality used for partitioning the superposed SCCs and to introduce the algorithmic and the functional relational constraints.

Additionally, we note that the OF was developed in order to offer some desired properties: precision of detection, robustness, coherence and stability. The OF is composed of three terms: the *data driven* E_D , the *model driven* E_M , and the smoothing E_S . E_D is a consistency measurement between the solution and the observed image. The form of E_D assures the coherent coexistence between the SCCs combined with the precision and the robustness properties of the discontinuity detection. E_M was chosen as a function of the thresholded statistical distribution of the features of each curve segment where curve segmentation graph is the dual of the region adjacency (relational) graph of the model. Additionally, the relational graph was used continuously to introduce algorithmic constraints in the deformation process in order to preserve the topological aspect of the solution. The function E_S was chosen as a weighted sum of the curve length and, the first and the second order quadratic variations of the local direction. E_S introduces a stability effect in the search technique with respect to noise.

We developed a search method with the following properties: (i) Move near the global optimum by using the global alignment module and a multiresolution module that also accelerate the convergence; (ii) Reduce the search space by using a representation of

deformable SCCs. In fact, the preservation of the SCCs bypasses the problem of edge organization and region merging; (iii) Enhance the search space convexity by using a far range and randomized deformation scheme in a pyramidal structure.

Experiments with noisy synthetic images and noisy X-ray images of a synthetic complex phantom gave satisfactory results. The detection was precise, robust and stable. Labeling and multiple SCC coherence were excellent. Experiments also proved the capability and robustness of the approach to detect in images of variable contrast and background.

Finally, the statistical representation of a partitioned curve can be considered as a good unbiased source of high level constraint in the search process when each curve segment shows a specific feature nodality. However, a feature dispersion can soften the *model driven* constraints. We are currently investigating the usefulness of hardening the *model driven* constraint by using distributive mapping technique between two SSCs.

Acknowledgments

H. Mallouche was supported in part by the National Sciences and Engineering Research Council of Canada (NSERC), the "Fonds pour la formation de chercheurs et l'aide à la recherche" of Quebec (FCAR), the Canadian Council of Professional Engineers-North American Life, and the cooperation France-Quebec, under scholarship numbers 167653, 961252, CCPE/NAL95-96, and 143690B, respectively. Y. Goussard was supported in part by the NSERC of Canada under Research Grant OGP0137417. J.A. de Guise was supported in part by the NSERC.

References

- Ayache, N., Cohen, I., and Herlin, I. 1992. Medical Image Tracking. In *Active Vision*, Blake, A. and Yuille, A., editors, MIT Press: Cambridge, pp. 285-302.
- Bajscy, R., and Kovacic, S. 1989. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46: 1-21.
- Ballard, D.H. and Brown, C. M. 1982. *Computer Vision*. Prentice-Hall: Englewood Cliffs, Inc.
- Besag, J. 1989. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(2): 395-407.

- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36: 192-236.
- Besag, J. 1975. Statistical analysis of non-lattice data, *The Statistician*, 24(3): 179-195.
- Besl, P.J. 1988. Geometric Modeling and Computer Vision. *Proceedings of the IEEE*, 76(8): 936-958.
- Chou, P.D. and Brown, C.M. 1990. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4: 185-210.
- Cohen, I. and Cohen, L.D. (1996). A Hybrid Hyperquadric Model for 2-D and 3-D Data Fitting. *Computer Vision and Image Understanding*, 63(3): 527-541.
- Cohn, H. 1967. *Conformal Mapping on Riemann Surfaces*. McGraw-Hill: New York.
- Close, R. A. 1991. Anatomic 3-D Reconstruction of Vascular Geometry from Two Orthogonal Angiographic Image Sequences. In *Proc. Medical Imaging V*, SPIE, Vol. 1445, pp. 513-522.
- Derin, H. and Elliot, H. 1987. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(1): 39-55.
- Dickinson, S.J., Pentland, A.P., and Rosenfeld, A. 1992. 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2): 174-198.
- Dubuisson, M.-P. and Jain, A.K. 1995. Contour Extraction of Moving Objects in Complex Outdoor Scenes. *International Journal of Computer Vision*, 14: 83-105.
- Engelking, R., and Sieklucki, K. 1992. *Topology A Geometric Approach*. Berlin: Heldermann Verlag.
- Grimson, W.E.L. 1990. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press: Cambridge.
- Horand, R. and Skordas, T. 1989. Stereo Correspondence Through Feature Grouping and Maximal Cliques. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(11): 1168-1180.
- Jain, A. K., Zhong, Y., and Lakshmanan, S. 1996. Object Matching Using Deformable Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(3):267-278.
- Kass, M., Witkin, A., and Terzopoulos, D. 1988. Snakes: Active contour Models. *International Journal of Computer Vision*, 1(4): 321-331.
- Lai, K. F. and Chin, R. T. 1995. Deformable Contours: Modeling and Extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(11):1084-1090.
- Lavallée, S. and Szeliski, R. 1995. Recovering the position and orientation of free-form objects from image contours using 3D distance maps. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(4): 378-390.
- Lowe, D.G. 1985. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers: Boston.
- Lowe, D.G. 1987. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31: 355-395.
- Malladi, R., Sethian, J.A., and Vemuri, B.C. 1995. Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(2):

158-175.

- Mallouche, H., de Guise, J., and Goussard, Y. 1997a. Probabilistic Model of Multiple Dynamic Curve Matching for Semitransparent Scene of Nonrigid Objects. Submitted to *IEEE Trans. Pattern Analysis and Machine Intelligence*.
- Mallouche, H., de Guise, J., and Goussard, Y. 1997b. Dynamic Model for Global Alignment and Scaling. Submitted to *Computer Vision and Image Understanding*.
- Mallouche, H., de Guise, J., and Goussard, Y. 1995. Probabilistic Model of Multiple Dynamic Curve Matching for Semitransparent Scene. In *Proc. Vision Geometry IV Conf.*, San Diego, SPIE, Vol. 2573, pp. 148-157.
- Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Co: San Francisco.
- Marroquin, J., Mitter, S., and Poggio, T. 1987. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82(397): 76-89.
- Nastar, C. and Ayache, N. 1996. Frequency-Based Nonrigid Motion Analysis: Application to Four Dimensional Medical Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(11): pp. 1067-1079.
- McCafferty 1993. *Human and Machine Vision: Computing Perceptual Organization*. Ellis Horwood: New York.
- Staib, L.H. and Duncan, J. S. 1992. Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(11): 1061-1075.
- Pentland, A. and Horowitz, B. 1991. Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7): 730-742.
- Storvik, G. 1994. A Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(10): 976-986.
- Suetens, P., Smets, C., Verbeeck, G., and Oosterlinck, A. 1988. Knowledge-Based Blood Vessel Delineation on DSA Images. In *Proc. Medical Imaging II*, SPIE, Vol. 914, pp. 406-411.
- Tenenbaum, J.M. and Barrow, H.G. 1977. Experiments in Interpretation-Guided Segmentation. *Artificial Intelligence*, 8: 241-274.
- Terzopoulos, D. 1986. Image Analysis Using Multigrid Relaxation Methods. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(4): 413-424.
- Terzopoulos, D. 1987. On matching deformable models to images. In *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, DC, pp. 160-163.
- Terzopoulos, D., Platt, J., Barr, A., and Fleischer, K. 1987. Elastically Deformable Models. *Computer Graphics*, 21(4): 205-214.
- Witkin, A., Terzopoulos, D., and Kass, M. 1987. Signal matching through scale space. *International Journal of Computer Vision*, 1: 133-144.
- Whittle, P. 1963. Stochastic Processes in Several Dimensions. *Bulletin of the International Statistical Institute*, 40: 974-994.

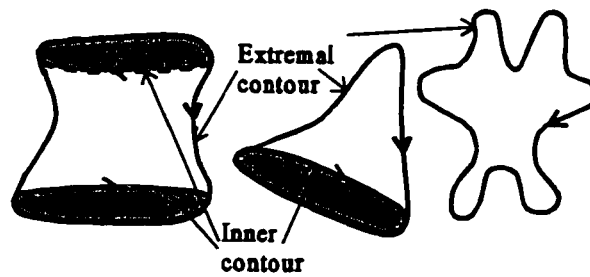


Fig. 5.1. Cylindric, conic and planar free forms. The 2-D semitransparent projection of these forms from a general viewing position can be described using one extremal curve with one or two inner curves.

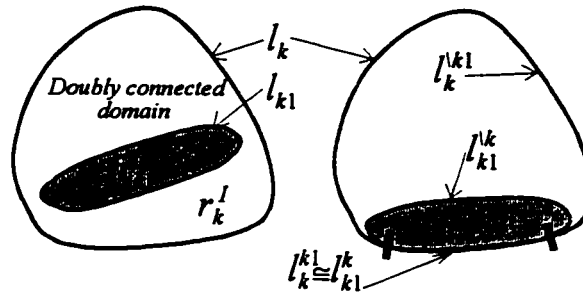


Fig. 5.2. Topologic relations between extremal curve l_k and inner curve l_{k1} . l_{k1} transforms the domain r_k from a *simply* connected region to *doubly* connected (left). Here, l_{k1} is considered as a hole in r_k . In the right scheme, $l_k \cap l_{k1} \neq \emptyset$. Hence, l_k and l_{k1} can be partitioned into one *simple* segment and one *double* segment, as $l_k = \{l_k^{k1}, l_k^k\}$ and $l_{k1} = \{l_{k1}^{k1}, l_{k1}^k\}$.

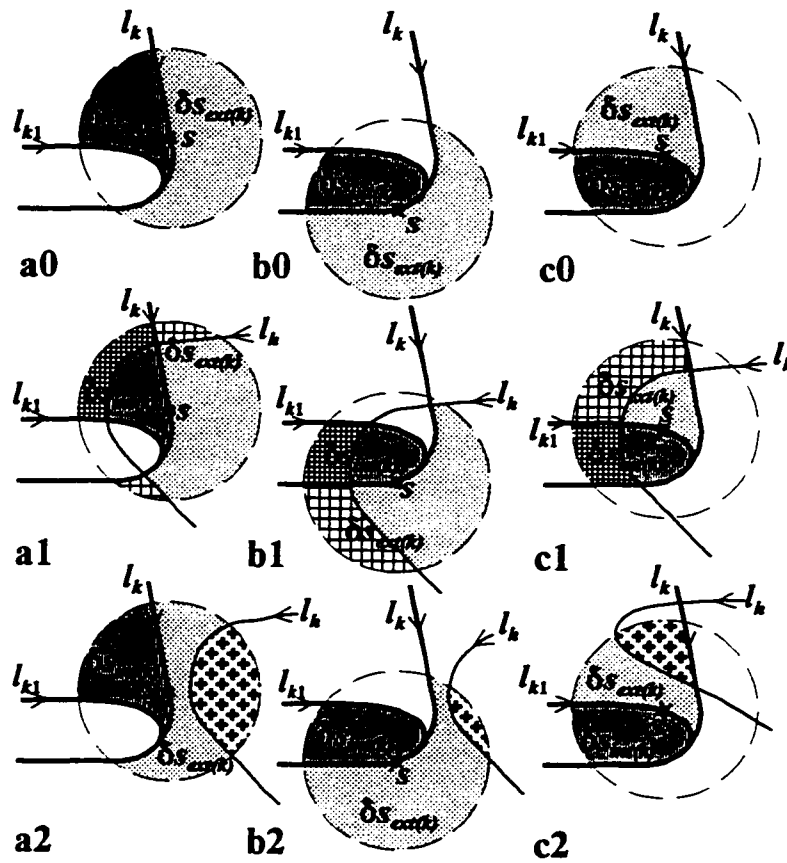


Fig. 5.3. Partition of the local neighborhood system δs for extremal-inner curves configurations. l_k , l_{kl} and l_h are *extremal*, *inner*, and *additional* curves, respectively. v_s is given by the union of the shaded parts of δs . When l_h does not cross l_k or l_{kl} (a2, b2 and c2) the sum of the weighting coefficients $\mu_s = \sum_{j=1, \dots, m} \mu_{s,j}$ equals to v_s minus the surface of the + shaded part.

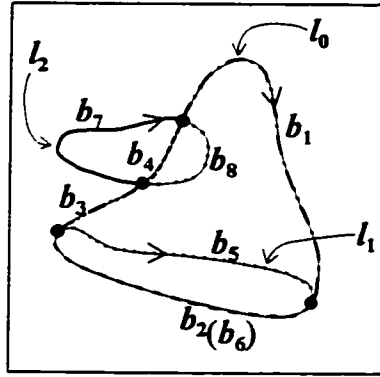


Fig. 5.4. Curve segmentation and labeling of the curve set $l = \{l_0, l_1, l_2\}$. Eight labeled segments are detected $l = \{l_0^{(b_1)}, l_0^{(b_2)}, l_0^{(b_3)}, l_0^{(b_4)}, l_1^{(b_5)}, l_1^{(b_6)}, l_2^{(b_7)}, l_2^{(b_8)}\}$.

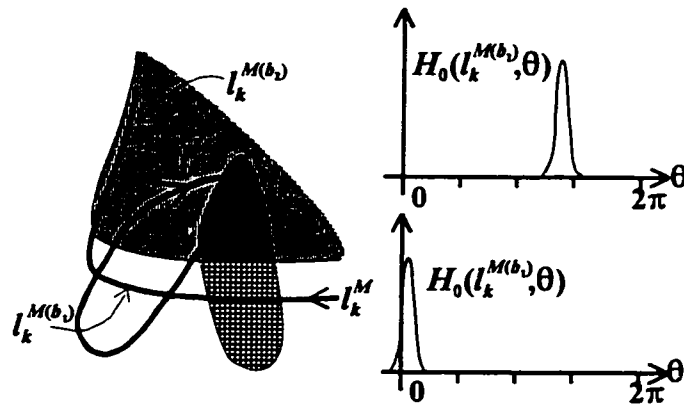


Fig. 5.5. Typical situation where the convergence is compromised by the *gluing* drawback the model driven term. The model and the intensity image are shown in solid lines and shaded regions, respectively. The curve l_h can not match exactly the real discontinuities because of the uninodality of $H_0(l_k^{M(b_2)}, \theta)$.

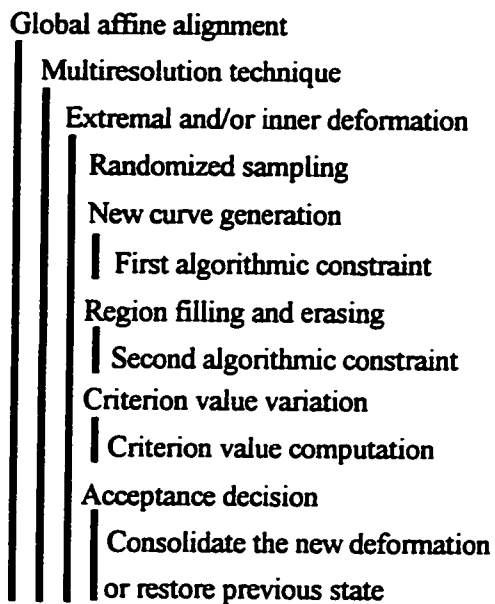


Fig. 5.6. The global search method.

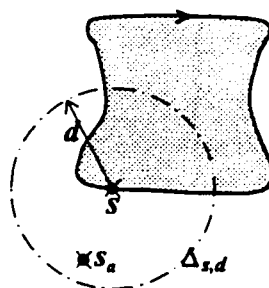


Fig. 5.7. Deformation technique with two sampling processes for s and s_a . The summit of the deformation, s_a , is sampled from uniform or intensity gradient distribution. s_a is confined to the maximum deformation disk $\Delta_{s,d}$.

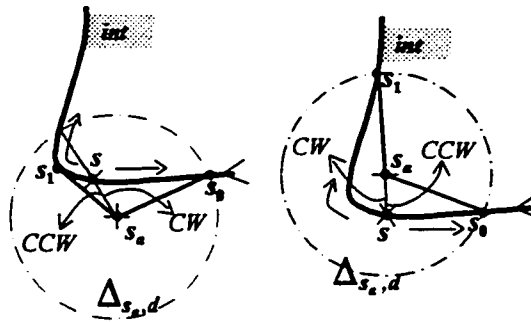


Fig. 5.8. Determination of the curve segment to deform $l_{k,(s_0,s_1)}$. The curve l_k is scanned *forward* and *backward* from s . $l_{k,(s_0,s_1)}$ is confined to the disk $\Delta_{s_a,d}$. The summit s_a can belong either to the outer region (left) or to the inner region (right) of l_k .

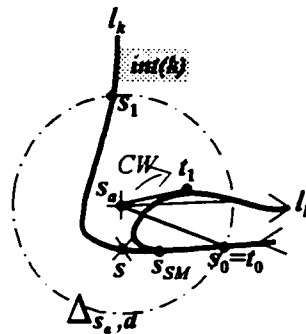


Fig. 5.9. Detection of the segment to deform $l_{h,(t_0,t_1)}$ of the *inner* curve l_h when the segment to deform $l_{k,(s_0,s_1)}$ of the *extremal* curve l_k contains a *splitting* point s_{SM} .

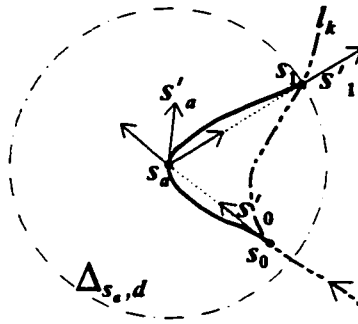


Fig. 5.10. Generation of the new segment $l_{k,(s_0,s_a,s_1)}$ using two *glued* segments of cubic spine $l_{k,(s_0,s_a)}$ and $l_{k,(s_a,s_1)}$.

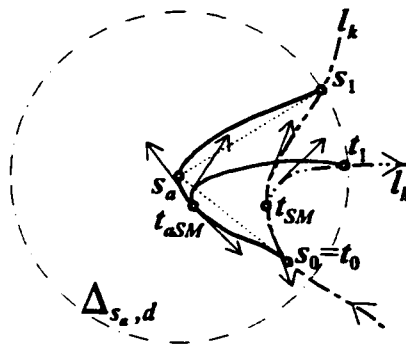


Fig. 5.11. Generation of the new *doubly* connected pair $l_{k,(s_0,s_a,s_1)}$ and $l_{h,(t_0,t_{aSM},t_1)}$. t_{SM} is the old *splitting* point and t_{aSM} is the generated *splitting* point.

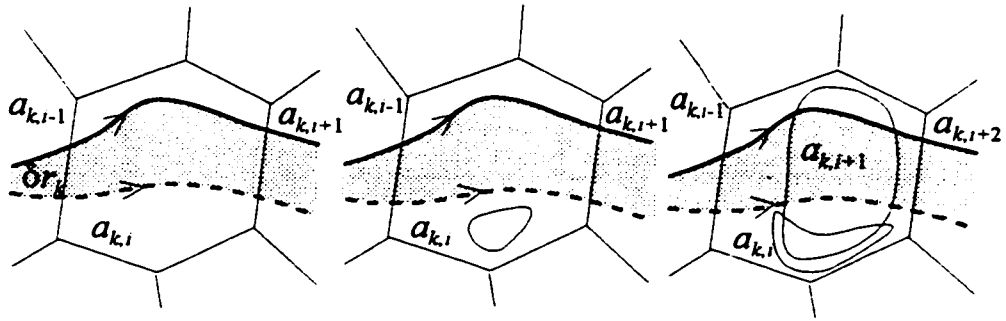


Fig. 5.12. Three schemes where the curve deformation $\delta l_k = l_{k,(s_0,s_1)} - l_{k,(s_0,...,s_1)}$ locally fulfill conditions 1 and 2. $l_{k,(s_0,s_1)}$: dashed line; $l_{k,(s_0,...,s_1)}$: solid line. Left: $r_k^{(a_{k,i})}$ is a *simply* connected region. Center and right: $r_k^{(a_{k,i})}$ is a *doubly* connected region enclosing a hole of one connected region (one simple closed curve) and a hole of three connected regions (two simple closed curves), respectively.



Fig. 5.13. Free conic (left) and free cylindric (right) object used to match with noisy synthetic images.

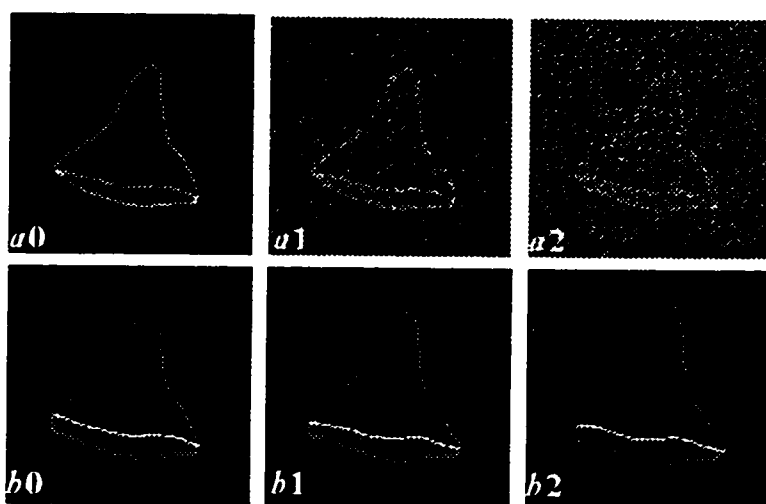


Fig. 5.14. Experiments with synthetic images modeling semitransparent projections of an arbitrary free conic object. Images a_1 and a_2 are generated by adding an i.i.d. $N(\mu=0; \sigma)$ noise of $\sigma=0.5$ and 1.0 to a_0 , respectively. b_0 , b_1 and b_2 are the corresponding matching results with a_0 , a_1 and a_2 , respectively.

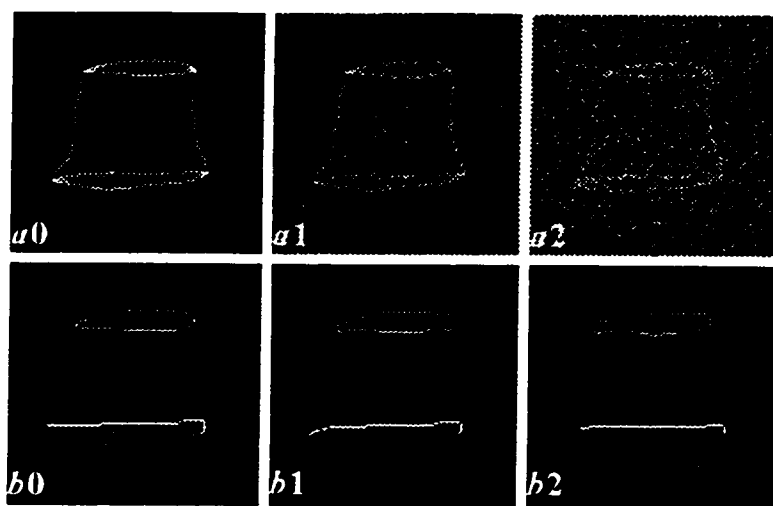


Fig. 5.15. Experiments with synthetic images modeling semitransparent projections of an arbitrary free cylindric form. Images and results are arranged in the same order of Fig. 5.14.

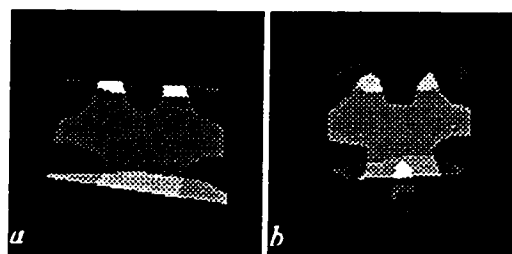


Fig. 5.16. Two hand-drawn wire-frame models to match with X-ray images.

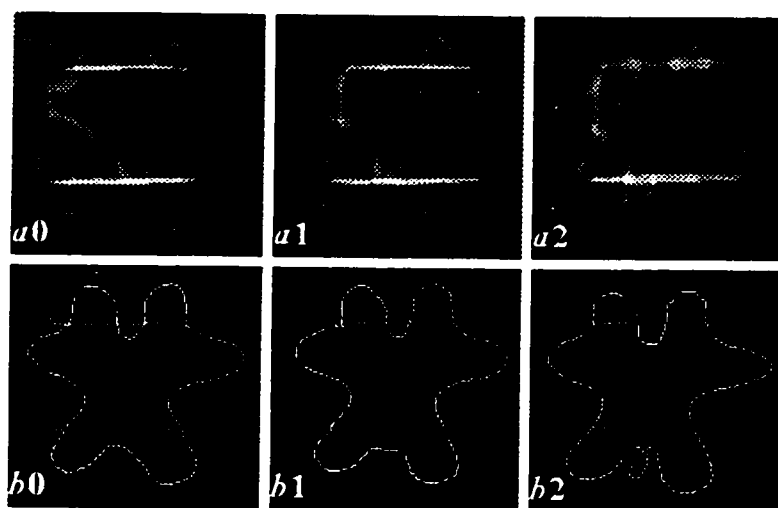


Fig. 5.17. Experiments with three X-ray images of the a synthetic object. Images a_0 , a_1 and a_2 have the same 2-D relational graph. b_0 , b_1 and b_2 are the matching results of a_0 , a_1 and a_2 with model a of Fig. 5.16 after 5 full sweep iterations, respectively.

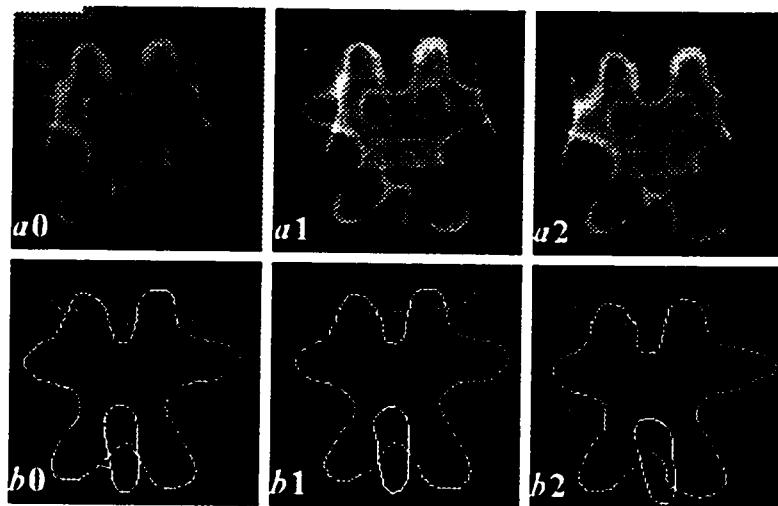


Fig. 5.18. Experiments with three X-ray images of the same object. Images a_0 , a_1 and a_2 have the same 2-D relational graph. Using the hand-drawn model b of Fig. 5.16, b_0 , b_1 and b_2 are the corresponding matching results after 5 full sweep iterations, with a_0 , a_1 and a_2 , respectively.

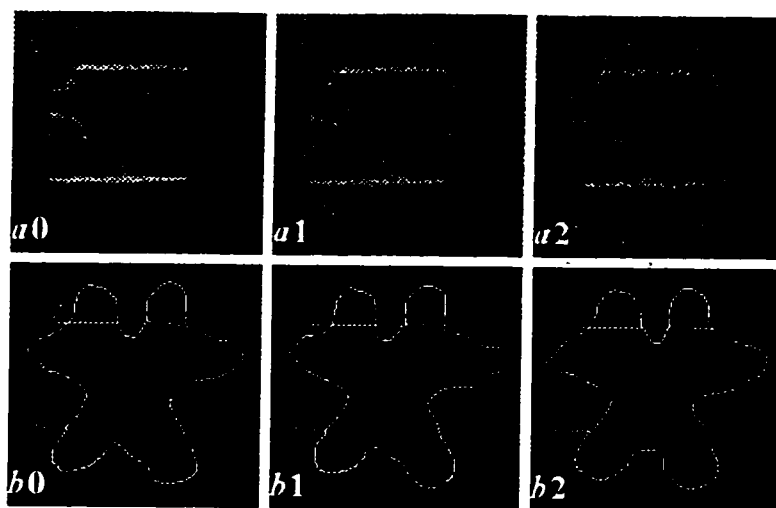


Fig. 5.19. Experiments with noise-corrupted images (see the text). b_0 , b_1 and b_2 are the corresponding matching results with a_0 , a_1 and a_2 , respectively.

Chapitre 6

Synthèse et discussion

Dans le présent chapitre, on récapitule et synthétise l'approche proposée dans cette thèse. La synthèse retrace les objectifs, les difficultés à surmonter, les hypothèses générales et résume la méthodologie adoptée. Nous discutons ensuite les limitations de l'approche et nous proposons de nouvelles voies pour surmonter ces limitations.

6.1 Synthèse

Nous avons développé dans cette thèse un système de vision artificielle pour la reconnaissance et la détection d'objets à partir d'images radiographiques. Le *tableau 6.1* présente une récapitulation générale des objectifs, des difficultés et des hypothèses sur lesquelles repose l'approche. Les objectifs indiqués par (R) résument les caractéristiques souhaitables de la méthode et l'espace de recherche. Concernant les difficultés, elles peuvent relever du processus de semitransparence de formation d'image (F) ou de la nature propre de la scène (N). L'approche a été fondée sur deux types d'hypothèses: les hypothèses communes (C) et les hypothèses simplificatrices (S). Les hypothèses communes relèvent de la modélisation (forme géométrique, structurelle) et de la représentation des objets. Les hypothèses

simplificatrices sont introduites afin de réduire partiellement la complexité du problème.

Tableau 6.1. Objectifs, difficultés et hypothèses.

<i>Objectifs</i>	<i>Difficultés</i>	<i>Hypothèses</i>
Délimiter l'objet et ses composants constituants.	(F1) Images à faible rapport S/B.	(C1) Appariement avec un modèle connu <i>a priori</i> .
Détecter les frontières avec précision et robustesse.	(F2) Images à contraste faible et variable.	(C2) Modèle déformable.
Étiqueter les régions.	(F3) Superposition des structures.	(C3) Non-accidentalité de la pose.
(R1) Réduire la dimension de l'espace de recherche.	(N1) Objets non rigides.	(C4) Objet à géométrie structurelle. Graphe relationnel intrinsèque.
(R2) Trouver une solution globale.	(N2) Surfaces d'objets à discontinuité réduite de premier ordre.	(C5) Préservation du contenu structurel durant la recherche.
	(N3) Objets à géométrie complexe.	(S1) Pose estimée par un processus externe.
	(N4) Fond variable et complexe.	(S2) Objet isolé.

L'approche proposée dans cette thèse se résume par les points du *tableau 6.2*. La représentation par MMC découle des hypothèses (C4), (C3) et (S1) et de l'objectif (R1). Les propriétés intrinsèques³⁴ des SCF aident à réduire considérablement l'espace de recherche. On a développé le critère de maximisation de la distribution de *pseudo-probabilité conjointe* pour une représentation *simple* de MMC (chapitre 4) et pour une représentation plus réaliste dite *complète* (chapitre 5) où en plus des courbes extrémales, on a inclus dans le modèle des

³⁴

La continuité, la fermeture et la dualité ligne-région.

courbes internes. Le critère sert de contrainte fonctionnelle pour guider le processus de recherche. La fonction d'objectif du critère s'exprime par la somme des interactions locales et est formée de trois termes résumés dans le *tableau 6.3*.

Tableau 6.2. Notre approche.

<i>Principe</i>	Reconnaissance, détection et étiquetage par appariement de modèle déformable avec des images de scènes semitransparentes.
<i>Domaine</i>	Dans le plan 2-D de l'image.
<i>Représentation</i>	Modèle de multiples SCF.
<i>Critère</i>	Probabiliste, global, à forme distribuée: <i>pseudo-probabilité conjointe</i> .
<i>Méthode de recherche descendante</i>	Alignement global du modèle. Structure pyramidale (multirésolution). Déformation aléatoire et progressive de la solution.

La méthode de recherche développée est descendante. Elle se fait en deux étapes: l'alignement global et la déformation progressive intégrée dans une structure pyramidale à multirésolution. En plus de la contrainte fonctionnelle, la recherche par déformation progressive de la solution est soumise à des contraintes algorithmiques qui vérifient les hypothèses de la *préservation de SCF* et de la *préservation du contenu relationnel du modèle*. Le *tableau 6.3* énumère toutes les contraintes. Celles indiquées par (T) relèvent des connaissances topologiques et statistiques du modèle.

Tableau 6.3. Les contraintes du processus de déformation progressive.

Type	Contrainte	Propriétés
Fonctionnelle	Terme relié aux données: mesure de similarité entre la solution et l'image observée.	Cohérence du MMC. Précision et robustesse de la détection. Faible répulsion.
	(T) Terme relié au modèle: mesure de similarité entre les attributs du graphe relationnel de la solution et celui du modèle.	Influence supérieure. Représentation statistique. Ne compromet pas la précision de la détection.
	Terme de lissage. Pénalise les variations quadratiques de l'orientation locale.	Premier ordre et deuxième ordre. Stabilisateur.
Algorithmique	Préservation des simples courbes fermées.	Implicite.
	(T) Préservation du graphe relationnel. 1- Préservation de la description séquentielle du segment déformé.	Vérification séquentielle durant le processus de déformation.
	2- L'ensemble des étiquettes des régions touchées est inclus dans celui du segment déformé.	Vérification pendant la mise à jour des régions.

L'appariement par déformation progressive est précédé d'une étape d'alignement global. Le but de cette étape est de rapprocher la solution initiale le plus possible de la solution finale avant le commencement de l'étape de déformation. Le *tableau 6.4* résume les grandes lignes de la technique d'alignement global. L'alignement se fait par transformation affine appliquée au modèle. Cette transformation est décomposée en une translation, une rotation, une mise à l'échelle et un cisaillement. La technique s'applique à différents types de modèles et d'images. Elle converge très bien même dans des conditions de grande disparité entre le modèle et l'image. Elle est également stable, robuste et précise.

Tableau 6.4. Alignement global du modèle.

<i>Principe</i>	Modélisation dynamique pour alignement global: translation, rotation, mise à l'échelle, et cisaillement
<i>Domaine</i>	2-D ou 3-D
<i>Représentation</i>	Intensité, segments de lignes, courbes, surfaces
<i>Critère</i>	Probabiliste, global, à forme distribuée: distribution de la probabilité de <i>vraisemblance</i>
<i>Méthode de recherche</i>	Descendante. Modélisation mécanique avec un système d'équations de mouvement dynamique. Mouvements hiérarchiques.

La technique de déformation aléatoire a été appliquée sur des MMC *simples* (chapitre 4) et des MMC *complets* (chapitre 5), et elle s'est révélée efficace pour appariement avec des images de scènes semitransparentes. Les expériences avec des images synthétiques fortement bruitées ont révélé la stabilité, la robustesse et la précision de la technique. D'autres expériences avec d'images rayon-X d'un fantôme de vertèbre à géométrie complexe ont confirmé ces propriétés et ont montré aussi la capacité de l'approche de détecter avec précision les éléments de l'objet dans les images bruitées à contraste variable.

Les contributions majeures de cette thèse sont énumérées au *tableau 6.5*. Contrairement aux méthodes basées sur les primitives (Lavallée et Szeiliski, 1995; Feldmar et Ayache, 1996) ou les descripteurs (Reeves *et al.*, 1988; Staib et Duncan, 1992; Wallace et Wintz, 1980), notre méthode d'alignement global ne nécessite pas une étape préliminaire d'extraction et regroupement, alors que par rapport à d'autres méthodes basées sur l'intensité (Nastar et Ayache, 1996; Pentland et Horowitz, 1991; Viola, 1995), notre méthode est peu sensible au bruit et aux différents paramètres du processus de formation d'image. Contrairement aux

modèles élastiques (Bajscy et Kovacic, 1989; Terzopoulos, Platt *et al.*, 1987), notre approche d'appariement introduit les contraintes reliées au modèle sans compromettre la précision de la détection même dans les situations de grandes disparités. Cela est dû à la forme libre et non paramétrique du modèle géométrique. De plus, notre approche ne dépend pas des variations du processus de formation d'image, contrairement aux méthodes d'appariement basée sur l'intensité qui nécessitent des modèles d'intensité (Amit *et al.*, 1991; Brunelli et Poggio, 1995). En effet, le modèle géométrique proposé dans cette thèse est une représentation de frontières, nous permettant d'éviter d'introduire un processus additionnel d'intensité pour la restauration d'image. Comparé à d'autres modèles de contours dynamiques (Lai et Chin, 1995; Storvik, 1994; Staib et Duncan, 1992; Geiger *et al.*, 1995), qui se sont limités à l'utilisation d'un seul contour ou des contours disjoints, notre modèle représente le processus de semitransparence en superposant des multiples SCF dans un cadre stochastique assurant la coexistence harmonieuse des courbes et utilisant les connaissances relationnelles du modèle

Tableau 6.5. Les contributions de la thèse.

<i>Alignement global</i>	<ul style="list-style-type: none"> - Nouvelle technique dans un cadre probabiliste avec un critère global. - Technique s'appliquant à plusieurs types de représentation. - Capacité améliorée de convergence même dans des conditions de grande disparité, à l'aide d'un modèle dynamique de transformations hiérarchiques. Technique robuste et stable.
<i>Modèle probabiliste</i>	<ul style="list-style-type: none"> - Nouveau modèle géométrique (<i>simple</i> et <i>complet</i>) à multiples simples courbes fermées. - Modélisation des scènes semitransparentes par la superposition des simples courbes fermées. - Distribution de <i>pseudo-probabilité conjointe</i> à forme distribuée intégrant des mesures d'adéquations entre la solution d'un côté et l'image et le modèle d'un autre côté. - Fonction d'objectif assurant la cohérence, la précision et la robustesse de la recherche. - Interactions entre la solution et les connaissances topologiques du modèle sans compromettre la précision de la détection.
<i>Méthode de recherche</i>	<ul style="list-style-type: none"> - Recherche par déformation aléatoire et progressive de la solution. - Transition aléatoire à échelle variable. - Préservation de la représentation en SCF et préservation du contenu relationnel du modèle.

pour contraindre le processus de recherche.

6.2 Voies futures de recherche

Dans cette section, il sera question des limitations de l'approche développée dans cette thèse et de voies futures qui, croyons-nous, aideront à surmonter ces limitations. Les deux premières limitations découlent directement des hypothèses simplificatrices (S1) de la pose connue approximativement et (S2) de l'objet isolé. La troisième limitation découle de la représentation statistique impliquée dans le terme relié au modèle. Nous avons introduit les

hypothèses simplificatrices dans le but de réduire la complexité du problème. Cependant, l'estimation initiale de la pose est souvent très approximative³⁵. De plus, si nous voulons appliquer notre approche à des images de rayon-X très complexes comme, par exemple, les radiographies de la colonne vertébrale, l'hypothèse de l'objet isolé ne tient plus à cause de l'imbrication des vertèbres.

Afin de réduire l'importance de ces deux hypothèses dans l'approche, on propose quatre voies futures de recherche:

1. intégrer la représentation du modèle dans un graphe d'aspect stochastique (Dickinson *et al.*, 1992, 1990 ; Grenander et Miller, 1994). Cela est désigné aussi par le terme de représentation reliée à l'observateur³⁶;
2. introduire dans le terme relié au modèle dans le critère de MPPC une mesure d'adéquation pour la préservation du squelette de l'objet à détecter (Ge et Fitzpatrick, 1996);
3. remplacer le terme relié au modèle par une mesure d'adéquation basée sur l'appariement point-à-point entre le modèle et la solution³⁷ (Mokhtarian et Mackworth, 1992; Witkin, 1983; Parent et Zucker, 1989);
4. intégrer l'ensemble dans un cadre évidentiel (Lowe, 1987; Chou et Brown, 1990; Charniak, 1983).

Ces directions de recherche seront élaborées par la suite et nous croyons qu'elles constituent des suites logiques à notre travail.

³⁵ L'erreur relative à l'estimation initiale de la pose peut être élevée.

³⁶ "Viewer centered" par rapport à "object centered."

³⁷ En opposition avec l'échantillonnage statistique utilisé dans la présente thèse.

6.2.1 Représentation reliée à l'observateur

Afin de tenir compte de l'erreur relative à l'estimation de la pose et d'assouplir l'effet contraignant de l'hypothèse de préservation de l'aspect topologique de la solution, on peut penser à un système qui propose plusieurs solutions initiales à contenus topologiques légèrement voisins. Pour chaque solution initiale, la recherche par appariement est répétée suivant la méthode développée dans cette thèse. Parmi les différentes solutions trouvées, la solution finale sera celle qui optimise le plus le critère d'appariement. Cela est pratiquement réalisable lorsque le nombre de solutions initiales proposés est relativement limité. Dans le cas contraire comme, par exemple, le cas de l'imbrication des objets, cette procédure risque d'être coûteuse en terme de temps de calcul.

Une autre voie possible serait d'intégrer l'ensemble de tous les graphes 2-D générés par la projection du modèle à partir de toutes les poses possibles dans un graphe d'aspect (Dickinson *et al.*, 1992, 1990). Un noeud et une branche du graphe d'aspect représentent respectivement le modèle topologique associé à une pose spécifique et une possibilité de transition entre deux poses voisines. La *figure 6.1* montre les graphes de deux projections 2-D issues de deux poses voisines. Il s'agit

d'introduire dans la méthode de recherche un processus de saut agissant en même temps que le processus de déformation (Grenander et Miller, 1994; Kimia *et al.*, 1995). Un saut représente un changement dans l'aspect topologique de la solution ou, de façon

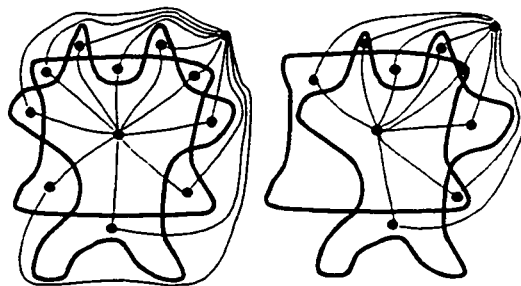


Figure 6.1. Graphes 2-D de deux projections issues à partir de deux directions voisines.

équivalente, un passage d'un noeud à un autre dans le graphe d'aspect (Dickinson *et al.*, 1992, 1990). Cette procédure enrichit sûrement la fonction de transition de la méthode de recherche en réduisant les contraintes algorithmiques et de ce fait augmente la dimension de l'espace de recherche.

6.2.2 La préservation du squelette

Dans la vision humaine, il est connu que l'information dérivant du squelette d'un objet quelconque est importante dans le processus de reconnaissance d'objet. Cette information structurelle peut être également utile en vision artificielle (Marr, 1982) et plus particulièrement, dans la reconnaissance des objets déformables ou articulés (Zhu et Yuille, 1996; Kimia *et al.*, 1995; Subirana-Vilanova, 1993; Ballard et Brown, 1982; Marr 1982). D'ailleurs, notre approche utilise aussi une partie du contenu informationnel de la structure du modèle pour contraindre la recherche de la solution. Cependant, l'information provenant du squelette du modèle n'est pas utilisée dans notre approche, ce qui peut dans certains cas causer une déstructuration du squelette de la

solution. La *figure 6.2* illustre un exemple d'amputation de forme. Cela peut arriver principalement au début du processus de recherche où le critère est encore loin de la bonne solution et la valeur du rayon maximal de déformation est encore élevée. Une déstructuration peut être due, par exemple, à

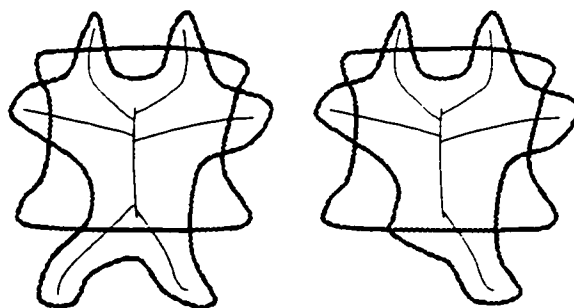


Figure 6.2. L'intégration d'une mesure de similarité de squelette dans une structure hiérarchique réduit la probabilité de déstructuration de la solution.

une amputation qui est un raccourci optimal vérifiant les contraintes algorithmiques et entraînant la solution dans un optimum local. Par conséquent, nous proposons d'inclure dans notre approche une procédure de mesure de correspondance entre le squelette du modèle et celui de la solution. Cette procédure nécessite évidemment que chaque déformation de la solution soit suivie d'une opération d'extraction du squelette puis d'une opération d'appariement entre les squelettes. Plusieurs techniques d'extractions de squelette ont été proposées (Zhu et Yuille, 1996; Ge et Fitzpatrick, 1996; Subirana-Vilanova, 1993). La limitation de cette voie demeure le temps de calcul élevé de l'opération d'extraction de squelette.

6.2.3 Mesure d'adéquation basée sur l'appariement point-à-point

On propose ici d'intégrer dans l'approche une mesure de similarité point-à-point entre la solution et le modèle utilisant une représentation de frontières (MMC) semblable à celle définie dans cette thèse. Le problème, tel que discuté dans l'article 3, sera de trouver une bijection entre la solution et le modèle. Le troisième terme du critère de MPPC régissant la technique de déformation intégrera ainsi cette mesure³⁸. L'avantage de cette approche par rapport aux modèles élastiques est que la déformation n'est pas limitée par l'élasticité de la matière. En effet, la similarité peut s'exprimer par une mesure de distance entre les descripteurs locaux³⁹ des points correspondants de la solution et du modèle sans tenir compte

³⁸ Cette mesure peut aussi représenter des modèles élastiques si elle est donnée par une distance euclidienne (Bajscy et Kovacic, 1989; Terzopoulos, Platt *et al.*, 1987).

³⁹ Par exemple, le vecteur de dérivées curvilignes de la courbe à un point donné (article 3). Une description locale à multi-échelle peut aussi être considérée (Mokhtarian et Mackworth, 1992; Witkin, 1983).

de la proximité de ces points.

Plusieurs techniques de correspondance entre deux courbes ont été proposées pour différents types d'applications comme, par exemple, la description (Mokhtarian et Mackworth, 1992), l'apprentissage et la reconnaissance (Ueda et Suzuki, 1993). Ces techniques reposent principalement sur deux opérations: la description multi-échelle des points de déflexion des courbes⁴⁰ (Witkin, 1983); et l'appariement entre les points de déflexion issus des différentes échelles de description (Fischler et Bolles, 1986; Lu et Jain, 1992; Ueda et Suzuki, 1993). Toutefois, ces techniques sont limitées par deux hypothèses: l'origine de balayage est connue et fixe; la vitesse de balayage curviligne est constante.

L'introduction de ces techniques dans notre approche peut aussi nous permettre d'éviter d'utiliser la mesure de similarité entre les squelettes. Cependant, l'inconvénient majeur de ces deux voies demeure le temps de calcul. De plus, il serait aussi intéressant d'explorer les techniques de mesure de correspondance utilisant la représentation par descripteurs de Fourier (Reeves *et al.*, 1988; Wallace et Wintz, 1980).

6.2.4 La recherche par raisonnement évidentielle

La quatrième voie propose d'intégrer l'ensemble de l'approche dans un cadre de recherche évidentielle (Lowe, 1987; Chou et Brown, 1990; Charniak, 1983). Cela consiste à explorer l'espace de recherche par étapes suivant une procédure hiérarchique. Le processus de recherche commence progressivement à appairier les parties du modèle donnant la

⁴⁰ Un point de déflexion sur une courbe 2-D est une position de changement de courbure entre un segment concave et un autre convexe. Donc, c'est un point à courbure nulle.

vraisemblance la plus élevée. La recherche se poursuit par ordre décroissant de mesure de confiance⁴¹. Cette démarche détecte d'abord les parties les plus évidentes ou les plus faciles à reconnaître. On utilise ensuite les parties déjà reconnues afin d'améliorer l'estimation de la probabilité *a posteriori* des éléments qui restent à découvrir. Pour les modèles déformables, cela peut inclure l'application d'une transformation géométrique de déformation, dont les paramètres sont déduits des parties détectées, sur les parties du modèle qui restent à apparier.

6.3 Discussions et recommandations

Les voies d'extension proposées dans la section 6.2 agissent sur différents aspects de notre approche afin d'améliorer ses performances⁴², sa robustesse et d'élargir ses applications. Dans ce sens, certaines hypothèses restrictives peuvent être assouplies permettant ainsi l'application de l'approche à des radiographies de la colonne vertébrale. Cependant, cette orientation implique l'assouplissement des contraintes de la méthode de recherche et par conséquent, pourrait compromettre les performances de la méthode. Pour les systèmes de reconnaissance d'objets, deux directions d'amélioration semblent possibles: assouplir les hypothèses et durcir les contraintes organisationnelles dérivant habituellement de la représentation (géométrique, relationnelle, etc) du modèle.

Ainsi, les voies proposées dans ce chapitre agissent sur la *représentation* des objets à reconnaître et sur la méthode de *recherche* (d'appariement). La méthode de recherche est

⁴¹ "Highest confidence first (Chou et Brown, 1990)."

⁴² La performance désigne ici la capacité de trouver la bonne solution.

généralement formée d'une *stratégie*⁴³ et des *contraintes* de recherche⁴⁴.

La voie 1 (représentation centrée sur l'observateur) enrichit la représentation relationnelle du modèle en utilisant un graphe d'aspect. Si cette voie n'est pas accompagnée d'un durcissement dans la contrainte *fonctionnelle* de recherche, elle peut entraîner une réduction des performances causée par l'assouplissement des contraintes *algorithmiques*. De plus, la progression dans cette voie nécessite la définition des modalités de transitions⁴⁵ inter-nodales du graphe d'aspect. Ces modalités doivent imposer des contraintes prévenant la déstructuration progressive de la solution pouvant se produire suite à une série de transitions inter-nodales dans le processus de recherche.

La voie 2 (préservation du squelette) et la voie 3 (appariement point-à-point) agissent sur la représentation géométrique du modèle et la contrainte fonctionnelle de la méthode de recherche. La voie 2 introduit le squelette dans le modèle géométrique. La voie 3 peut inclure une représentation multi-échelle dans les descripteurs locaux utilisés dans la mesure de similarité point-à-point (article 3). Le squelette peut représenter l'information extraite d'une échelle grossière réduisant les détails et conservant la structure de l'objet. En ce sens, les deux voies, 2 et 3, ajoutent à la représentation géométrique de l'objet une information structurelle. De plus, ces deux voies modifient le terme relié au modèle en durcissant la contrainte fonctionnelle qui en découle. Pour la voie 2, il faut ainsi développer une mesure de similarité entre les squelettes de la solution et du modèle sans compromettre la précision de la détection, et déterminer la technique d'extraction de squelette la plus efficace. En ce qui concerne la

⁴³ Dans notre cas, elle est descendante et à déformation progressive.

⁴⁴ Dans notre cas, une contrainte *fonctionnelle* et des contraintes *algorithmiques*.

⁴⁵ Une transition permet la modification du contenu topologique de la solution.

similarité des courbes (voie 3), on doit trouver une mesure⁴⁶ qui préserve la précision de la détection et développer la fonction de correspondance point-à-point⁴⁷ entre la solution et le modèle.

Quant à la voie 4 (raisonnement évidentiel), elle agit uniquement sur la stratégie de recherche affectant directement les performances de l'appariement. La concrétisation du paradigme évidentiel et l'intégration de notre approche à celle-ci restent à déterminer.

Malgré les bonnes performances de l'approche proposée dans cette thèse lorsqu'elle est appliquée à des images bruitées ayant un contraste faible et variable, il faut aussi s'intéresser au problème de l'occlusion dans les images de radiographie médicales. Lorsque certaines structures sont occluses dans l'image, les parties correspondantes de la solution risquent de converger vers des structures voisines plus visibles. Une façon de surmonter ce problème est d'utiliser des modèles élastiques dont les limitations ont été largement discutées dans cette thèse. Nous croyons qu'une représentation libre, telle que proposée dans les voies 2 et 3, combinée à l'approche développée dans cette thèse et intégrée à une stratégie de recherche évidentielle (voie 4) permettra de réduire la sévérité du problème d'occlusion tout en évitant les limitations des modèles élastiques.

Finalement, les recommandations découlant de cette discussion sont les suivantes:

1. préserver la nature libre du modèle afin d'assurer une précision élevée dans la détection des discontinuités;

⁴⁶ Les représentations de squelette et de MMC peuvent utiliser la même mesure de similarité car les deux sont de modèles à lignes.

⁴⁷ Plusieurs méthodes de correspondance utilisent une représentation multi-échelles et un balayage curviligne. Ici, deux questions se présentent: comment déterminer les origines des courbes et faut-il opter pour une vitesse de balayage constante ou variable?

2. améliorer la représentation du modèle en suivant les voies 2 et 3;
3. rehausser le terme relié au modèle en suivant les voies 2 et 3;
4. améliorer la stratégie de recherche en l'intégrant dans la voie 4;
5. si les expériences avec des images de scènes semitransparentes synthétiques composées d'objets imbriqués⁴⁸ de même que les expériences avec images radiographiques de la colonne vertébrale sont concluantes, aborder l'assouplissement de l'hypothèse (S1) en suivant la voie 1.

⁴⁸

Nous croyons que les voies 2 et 3 combinées à la voie 4 permettront d'assouplir l'hypothèse (S2) d'objet isolé.

Conclusion

Nous avons présenté dans cette thèse une nouvelle approche de segmentation et d'étiquetage d'images de rayon-X. Le système développé s'applique aux images d'objets non-rigides dérivant d'un processus semitransparent de formation d'image. Le système est particulièrement voué à la reconnaissance de structures complexes d'images radiographiques (par exemple, les images radiographiques de la colonne vertébrale). Notre approche est basée sur un modèle. Le modèle a représenté l'ensemble des informations connues *a priori* de la nature de la scène et ses objets et du processus de formation d'image. L'approche a consisté à apparier un modèle géométrique avec l'image observée. Notre système a utilisé une représentation géométrique basée sur les frontières convenant à la modélisation des projections semitransparentes de structures osseuses car l'information visuelle contenue dans les images radiographiques est plus existante au niveau des frontières des régions. De plus, il est difficile de modéliser l'intensité de l'image à cause de l'hétérogénéité de l'atténuation photonique des tissus, de la superposition des structures et de l'arrière-plan variable de l'image.

Au chapitre 1, nous avons énoncé les objectifs et on a posé la problématique générale de la vision artificielle et la problématique particulière aux scènes semitransparentes. Nous avons aussi présenté la problématique relevant aux applications de la vision artificielle en imagerie médicale et plus précisément, dans le contexte de la reconnaissance et la détection des objets

à partir d'images radiographiques. Nous avons ensuite émis les hypothèses sur lesquelles notre système est basé.

Le modèle 2-D est formé d'un ensemble de simples courbes fermées. Ces courbes possèdent des propriétés intrinsèques intéressantes qui intègrent certaines règles d'organisation perceptuelle. La superposition de ces courbes représente l'ensemble des frontières de la projection semitransparente de l'objet à partir d'une pose non-accidentelle. Nous avons relié cette représentation 2-D à la géométrie 3-D structurelle de l'objet. Un objet peut ainsi être constitué d'un assemblage de sous-objets élémentaires et simples suivant un graphe 3-D inhérent à l'objet. Nous avons choisi d'utiliser trois sortes de sous-objets libres de formes planes, coniques et cylindriques. De cette façon, une projection 2-D d'un objet a été représentée par un ensemble de simples courbes fermées extrémales et de courbes internes.

L'étiquetage des régions et la détection des discontinuités se sont fait par appariement du modèle avec l'image observée en deux étapes: l'alignement global puis l'appariement par déformation du modèle. L'approche globale a été décrite au chapitre 2.

Au chapitre 3, nous avons développé une nouvelle technique d'alignement global basée sur l'optimisation d'une fonction d'objectif. Celle-ci est issue du critère de *maximisation de vraisemblance*. Cette technique sert à aligner et à ajuster le modèle à l'image avant de commencer l'étape d'appariement. L'expression distribuée du critère se traduit par un champ de force dont les résultantes alignent le modèle à l'image en utilisant une structure hiérarchique de transformations linéaires: translation, rotation et mise à l'échelle dans deux directions non orthogonales. La technique a ensuite été testée avec différents types de

modèles (de frontières, d'intensité) et d'images (synthétiques, de rayon-X, photographiques). Elle a démontré d'excellentes propriétés de convergence, de stabilité et de robustesse.

Le modèle transformé par alignement global est par la suite déformé pour détecter les discontinuités d'intensité dans l'image. Cette deuxième étape a été décrite aux chapitres 4 et 5. Nous avons choisi de guider le processus de déformation par un critère global et un ensemble de contraintes algorithmiques. Le critère global de la distribution de *pseudo-probabilité conjointe* a été développé au chapitre 3 pour un modèle *simple* de multiples simples courbes fermées. Le critère possède des propriétés intrinsèques intéressantes: la forme distributive, le pouvoir d'intégrer la précision, la robustesse et la mesure locale de cohérence. La fonction d'objectif profite de la dualité ligne-région des SCF pour introduire une interaction entre la solution et les informations topologiques du modèle exprimées par des distributions statistiques des paramètres locaux. Le modèle *simple* de multiples SCF a ensuite été appliqué avec succès sur des images synthétiques fortement bruitées.

Le chapitre 5 a été consacré au modèle *complet* de multiples SCF. Le critère global a été modifié afin de pouvoir représenter le modèle *complet*. La suite du chapitre a été centrée sur l'aspect topologique du modèle et de la méthode de recherche. Cette méthode utilise une technique progressive de déformation aléatoire. L'acceptation d'une déformation proposée est régie par une contrainte fonctionnelle exprimée par l'optimisation de la fonction d'objectif et des contraintes algorithmiques imposant la préservation des SCF et de l'aspect topologique de la solution. Le modèle *complet* de multiples SCF a finalement été appliqué avec succès sur des images synthétiques bruitées et des images réelles d'un objet complexe.

L'approche, avec ses objectifs, hypothèses et propriétés a été récapitulée au chapitre 6.

Nous avons aussi discuté ses limitations et ses faiblesses. Nous avons ensuite proposé quatre voies pour surmonter ces limitations et assouplir les hypothèses simplificatrices. Cela permettra éventuellement d'augmenter les performances et la robustesse du système afin de pouvoir l'appliquer à des images radiographiques plus complexes. Nous soulignons que l'approche peut analyser des images moins compliquées que les radiographies de la colonne vertébrale comme, par exemple, les images issues des procédures automatisées d'inspection industrielle et de contrôle de qualité faisant appel à des processus semitransparents de formation d'images. Les performances de l'approche peuvent être améliorées par l'utilisation d'un cadre de recherche évidentiel (Lowe, 1987; Chou et Brown, 1990; Charniak, 1983) et d'une représentation plus descriptive et par conséquent plus contraignante sans affecter la précision de la détection. À ce niveau, nous croyons que l'approche améliorée suivant les recommandations proposées à la section 6.3 pourra être appliquée directement à des radiographies du rachis et par conséquent permettra d'incorporer les résultats obtenus dans la technique de reconstruction 3-D développée par notre groupe de recherche (de Guise *et al.*, 1995; Dansereau *et al.*, 1990; Marzan, 1976).

Les contributions de notre travail ont aussi été discutées au chapitre 6. Elles se résument par les points suivants:

1. développement d'une nouvelle technique d'alignement global basée sur un système dynamique de second ordre permettant la translation, la rotation et l'ajustement d'échelle dans deux directions non orthogonales. La technique est robuste, stable et s'applique sur différents types de représentations;
2. mise au point d'un nouveau modèle formé de la superposition de multiples simples

courbes fermées (extrémale et interne) pour représenter la semitransparence les projections radiographiques;

3. développement d'une fonction d'objectif basée sur la distribution *pseudo-jointe* et donnée par la somme d'un terme relié à l'image, d'un terme relié au modèle et d'un terme de lissage;
4. intégration d'informations relationnelles dans le terme relié au modèle sans compromettre la précision de la détection;
5. développement d'une méthode de recherche par déformation progressive et aléatoire guidée par une fonction d'objectif;
6. formulation de la préservation des informations relationnelles de la solution sous forme de contraintes algorithmiques.

Finalement, nous croyons que les travaux présentés dans cette thèse ont contribué à pousser les frontières de la recherche en vision artificielle et en imagerie médicale dans le cadre de l'analyse des scènes semitransparentes. Ce travail a aidé à comprendre davantage les problèmes de l'analyse des scènes semitransparentes et a tracé la base d'une nouvelle approche pour résoudre ces problèmes. Nous croyons aussi que ce travail aidera les futures recherches dans le développement de systèmes plus performants en vision artificielle.

Références

1. AMIT, Y., GRENANDER, U. et PICCIONI, M. (1991). Structural Image Restoration Through Deformable Templates. *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 376-387.
2. AUBIN, C.-E. (1995). Rôle biomécanique de la cage thoracique dans le traitement de la scoliose par orthèse. Thèse de doctorat, École polytechnique de Montréal.
3. AYACHE, N., COHEN, I. et HERLIN, I. (1992). Medical Image Tracking. In *Active Vision*, Blake, A. et Yuille, A., editors, Cambridge: MIT Press, pp. 285-302.
4. AYACHE, N. et FAUGERAS, O.D. (1986). HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 44-54.
5. BAJSCY, R. et KOVACIC, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, vol. 46, pp. 1-21.
6. BALLARD, D.H. et BROWN, C. M. (1982). *Computer Vision*. Englewood Cliffs: Prentice-Hall, Inc.
7. BARDINET, E., COHEN, L.D. et AYACHE, N. (1996). Analyzing the deformation of the left ventricle of the heart with a parametric deformable model. INRIA, Rapport de recherche no. 2797.
8. BARRETT, H.H. et SWINDELL, W. (1981). *Radiological Imaging: Volume 1-2*.

Academic Press.

9. BERTERO, M., POGGIO, T.A. et TORRE, V. (1988). Ill-Posed Problems in Early Vision. *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869-889.
10. BESAG, J. (1989). Towards Bayesian image analysis. *Journal of Applied Statistics*, vol. 16, no. 2, pp. 395-407.
11. BESAG, J. (1975). "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, pp. 179-195.
12. BESAG, J. (1975). Statistical analysis of non-lattice data, *The Statistician*, vol. 24, no. 3, pp. 179-195.
13. BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 192-236.
14. BESL, P.J. (1988). Geometric Modeling and Computer Vision. *Proceedings of the IEEE*, vol. 76, no. 8, pp. 936-958.
15. BESL, P.J. et JAIN, R.C. (1985). Three-Dimensional Object Recognition. *Computing Surveys*, vol. 17, no. 1, pp. 75-145.
16. BLAKE, A. et ZISSERMAN, A. (1987). *Visual Reconstruction*. Cambridge, MA: MIT Press.
17. BROOKS, R.A. (1981). Symbolic Reasoning Among 3-D Models and 2-D Images. *Artificial Intelligence*, vol. 17, no. 1-3, pp. 285-348.
18. BROOKS, R.A. (1986). Model-Based 3-D interpretation of 2-D Images. In *From Pixels to Predicates: Recent Advances in Computational and Robotis Vision*, Pentland, A.P., editor, Norwood: Ablex Publishing Co., pp. 299-321.

19. BRUNELLI, R. et POGGIO, T. (1995). Template Matching: Matched Spatial Filters and beyond. *Artificial Intelligence Laboratory*, MIT, A.I. Memo no. 1549.
20. CANNY, J. (1986). A Computational Approach to Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698.
21. CHARNIAK, E. (1983). The Bayesian Basis of Common Sense Medical Diagnosis. *Proceedings of the American Association for Artificial Intelligence*, Washington, D.C., August 1983, pp. 70-73.
22. CHOU, P.D. et BROWN, C.M. (1990). The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, vol. 4, pp. 185-210.
23. CLOSE, R. A., (1991). Anatomic 3-D Reconstruction of Vascular Geometry from Two Orthogonal Angiographic Image Sequences. *SPIE, Vol. 1445, Medical Imaging V*, pp. 513-522.
24. COHEN, I. et COHEN, L.D. (1996). A Hybrid Hyperquadric Model for 2-D and 3-D Data Fitting. *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 527-541.
25. COHEN, L. (1991). On active contour models and balloons. *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 53, no. 2, pp. 211-218.
26. COHN, H. (1967). *Conformal Mapping on Riemann Surfaces*. New York: McGraw-Hill.
27. CURWEN, R. et BLAKE, A. (1992). Dynamic Contours: Real-time Active Splines, in *Active Vision*, ed. A. Blake and A. Yuille. Cambridge: MIT Press, pp. 39-58.
28. DANSEREAU, J., BEAUCHAMP, A., DE GUISE, J.A. et LABELLE, H. (1990). Three-Dimensional Reconstruction of the Spine and the Rib Cage from

- Stereoradiographic Image Techniques. *Société Canadienne de G.M.*, vol. 2, pp. 61-64.
29. DANSEREAU, J. et STOKES, I.A.F. (1988). Measurements of the Three-Dimensional Shape of the Rib Cage. *Journal of Biomechanics*, vol. 21, no. 11, pp. 893-901.
 30. DECARLO, D. et METAXAS, D. (1996). Blended Deformable Models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 443-448.
 31. DE GUISE, J.A., MALLOUCHE, H., DANSEREAU, J. et LABELLE, H. (1995). Technique d'imagerie appliquées à la biomécanique rachidienne. *Rachis*, vol. 7, no. 3, pp. 135-144.
 32. DERIN, H. et ELLIOT, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39-55.
 33. DICKINSON, S.J., PENTLAND, A.P. et ROSENFELD, A. (1992). 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 174-198.
 34. DICKINSON, S.J., PENTLAND, A.P. et ROSENFELD, A. (1990). A representation for Qualitative 3-D Object Recognition Integrating Object-Centered and Viewer-Centered Models. In *Science of Vision*, Leibovic, K.N., editor. Springer-Verlag, pp. 398-421.
 35. DUBUISSON, M.-P. et JAIN, A.K. (1995). Contour Extraction of Moving Objects in Complex Outdoor Scenes. *International Journal of Computer Vision*, vol. 14, pp. 83-105.
 36. ENGELKING, R. et SIEKLUCKI, K. (1992). *Topology A Geometric Approach*.

Berlin: Heldermann Verlag.

37. FELDMAN, J.A. et YAKIMOVSKY, Y. (1974). Decision Theory and Artificial Intelligence: I. A Semantics-Based Region Analyser. *Artificial Intelligence*, vol. 5, pp. 349-371.
38. FELDMAN, J. et AYACHE N. (1996). Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, vol. 18, no. 2, pp. 99-119.
39. FISCHLER, M.A. et BOLLES, R.C. (1986). Perceptual Organization and Curve Partitioning. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 100-105.
40. FISHER, R. B., (1989). *From Surfaces to Objects: Computer Vision and 3-D Scene Analysis*. John Wiley & Sons.
41. GE, Y. et FITZPATRICK, J.M. (1996). On the Generation of Skeletons from Discrete Euclidean Distance Maps. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 11, pp. 1055-1066.
42. GEIGER, D., GUPTA, A., COSTA, L.A. et VLONTZOS, J. (1995). Dynamic Programming for Detecting, Tracking, and Matching Deformable Contours. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 294-302.
43. GEMAN, S. et GEMAN, D. (1985). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721-741.
44. GIBSON, J.J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.

45. GIBSON, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
46. GOLDBERG, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Co.
47. GORDON, I.E. (1989). *Theories of Visual Perception*. John Wiley and Sons.
48. GRENANDER, U. et MILLER, M.I. (1994). Representations of Knowledge in Complex Systems. *Journal of the Royal Statistical Society, Series B*, vol. 56, no. 4, pp. 549-603.
49. GRIMSON, W.E.L. (1981). *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press.
50. GRIMSON, W.E.L. (1990). *Object Recognition by Computer: The Role of Geometric Constraints*. Cambridge: MIT Press.
51. HAJEK, B. (1988). Cooling Schedules for Optimal Annealing. *Mathematics of Operations Research*, vol. 13, no. 2, pp. 311-329.
52. HOFFMAN, K.R., DOI, K., CHAN, H.-P. et TAKAMIYA, M. (1988). 3-D Reproduction of Coronary Vascular Trees Using the Double-Square-Box Method of Tracking. *SPIE, Vol. 914, Medical Imaging II*, pp. 375-378.
53. HOLLAND, J.H. (1975). *Adaptation of Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
54. HORAUD, R. et SKORDAS, T. (1989). Stereo Correspondence Through Feature Grouping and Maximal Cliques. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 11, pp. 1168-1180.

55. HORN, B.K.P. (1986). *Robot Vision*. Cambridge: The MIT Press.
56. HU, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, vol. IT-8, pp. 179-187.
57. HUTTENLOCHER, D.P. et ULLMAN, S. (1987). Object Recognition Using Alignment. *IEEE, First International Conference on Computer Vision*, London, pp. 102-111.
58. JACOBS, D.W. (1992). Recognizing 3-D Objects Using 2-D Images. *Ph.D. these, AIL, MIT: Massachusetts*.
59. JAIN, A.K., ZHONG, Y. et LAKSHMANAN, S. (1996). Object Matching Using Deformable Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 267-278.
60. KANADE, T. (1978). Region Segmentation: Signal v's Semantics. *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pp. 95-105.
61. KANADE, T. (1981). Recovery of the Three-dimensional Shape of an Object from a Single View. *Artificial Intelligence*, vol. 17, no. 1-3, pp. 409-460.
62. KAPUR, T. (1995). Segmentation of Brain Tissue from Magnetic Resonance Images. *Artificial Intelligence Laboratory, A.I. Technical Report No. 1566*.
63. KASS, M., WITKIN, A. et TERZOPOULOS, D. (1988). Snakes: Active contour Models. *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331.
64. KIMIA, B.B., TANNENBAUM, A.R. et ZUCKER, S.W., (1995). Shapes, Shocks, and Deformations I: The Components of Two-Dimensional Shape and the Reaction-Diffusion Space. *International Journal of Computer Vision*, vol. 15, pp.

189-224.

65. KIRKPATRICK, S., GELATT, C.D. et VECCHI, M.P. (1983). Optimization by Simulated Annealing. *Science*, vol. 220, no. 4598, pp. 671-680.
66. KOFFKA, K. (1935). *Principles of Gestalt Psychology*. Harcourt, Brace and Word, Inc.
67. LAI, K.F. et CHIN, R.T. (1995). Deformable Contours: Modeling and Extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1084-1090.
68. LAVALLÉE, S. et SZELISKI, R. (1995). Recovering the position and orientation of free-form objects from image contours using 3D distance maps. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 4, pp. 378-390.
69. LEYMARIE, F. et LEVINE, M.D. (1993). Tracking Deformable Objects in the Plane Using an Active Contour Model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 617-634.
70. LOWE, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31: 355-395.
71. LOWE, D.G. (1985). *Perceptual Organization and Visual Recognition*. Boston: Kluwer Academic Publishers.
72. LU, Y. et JAIN, R. (1992). Reasoning about Edges in Scale Space. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 450-468.
73. MALLADI, R., SETHIAN, J.A. et Vemuri, B.C. (1995). Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Trans. Pattern Analysis and Machine*

Intelligence, vol. 17, no. 2, pp. 158-175.

74. MALLOUCHE, H., DE GUISE, J. et GOUSSARD, Y. (1997a). Dynamic Model for Global Alignment and Scaling. Submitted to the *Computer Vision and Image Understanding*.
75. MALLOUCHE, H., DE GUISE, J. et GOUSSARD, Y. (1997b). Probabilistic Model of Multiple Dynamic Curve Matching for Semitransparent Scene of Nonrigid Objects. Submitted to the *IEEE Trans. Pattern Analysis and Machine*.
76. MALLOUCHE, H., GOUSSARD, Y. et DE GUISE, J. (1997c). Multiple Simple Closed Curves for Free Form Model Matching with Images from Semitransparent Scene. Submitted to the *International Journal of Computer Vision*.
77. MALLOUCHE, H., DE GUISE, J.A. et GOUSSARD, Y. (1995). Probabilistic Model of Multiple Dynamic Curve Matching for A Semitransparent Scene. In *Vision Geometry IV*, Metler, R.A., Wu, A.Y., Bookstein, F.L., Green, W.D.K., editors, Proc. SPIE 2573, pp.148-157.
78. MARR, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W. H. Freeman and Co.
79. MARROQUIN, J., MITTER, S. et POGGIO, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76-89.
80. MARZAN, G.T. (1976). *Rational design for close-range photogrammetry*, Ph.D. thesis, University of Illinois

81. MCCAFFERTY, J.D. (1993). *Human and Machine Vision: Computing Perceptual Organization*. New York: Ellis Horwood.
82. METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N. et TELLER, A.H. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092.
83. MODESTINO, J.W. et ZHANG, J. (1992). A Markov Random Field Model-Based Approach to Image Interpretation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 606-615.
84. MOHAN, R. et NEVATIA, R. (1992). Perceptual Organization for Scene Segmentation and Description. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 616-635.
85. MOKHTARIAN, F. et MACKWORTH, A.K. (1992). A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 789-805.
86. NASTAR, C. et AYACHE, N. (1996). Frequency-Based Nonrigid Motion Analysis: Application to Four Dimensional Medical Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 11, pp. 1067-1079.
87. O'SULLIVAN, F. et QIAN, M. (1994). A Regularized Contrast Statistic for Object Boundary Estimation-Implementation and Statistical Evaluation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 561-570.
88. PARENT, P. et ZUCKER, S.W. (1989). Trace Inference, Curvature Consistency, and Curve Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no.

- 8, pp. 823-839.
89. PENTLAND, A. et HOROWITZ, B. (1991). Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 730-742.
 90. POGGIO, T., TORRE, V. et KOCH, C. (1985). Computational Vision and Regularization Theory. *Nature*, vol. 317, pp. 314-319.
 91. PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A. et VETTERLING, W.T. (1988). *Numerical Recipes in C*. Cambridge University Press.
 92. REEVES, A.P., PROKOP, R.J., ANDREWS, S.E. et KUHL, F.P. (1988). Three-dimensional shape analysis using moments and Fourier descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 937-943.
 93. ROCK, I. (1975). *An Introduction to Perception*. New York: Macmillan Publishing Co.
 94. ROCK, I. (1983). *The Logic of perception*. The MIT Press.
 95. ROSENFELD, A. (1988). Computer Vision: Basic Principles. *Proceedings of the IEEE*, vol. 76, no. 8, pp. 863-868.
 96. SARKAR, S. et BOYER, K.L. (1993). Perceptual Organization in Computer Vision: A Review and a Proposal for a Classificatory Structure. *IEEE Trans. Systems, Man, and Cybernetics*, vol. 23, no. 2, pp. 382-399.
 97. SCHINZINGER, R. et LAURA, P.A.A. (1991). *Conformal Mapping: Methods and Applications*. Amsterdam: Elsevier.
 98. STAIB, L.H. et DUNCAN, J.S. (1992). Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14,

- no. 11, pp. 1061-1075.
99. STORVIK, G. (1994). A Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 976-986.
 100. SUBIRANA-VILANOVA, J.B. (1993). *Mid-Level Vision and Recognition of Non-Rigid Objects*, Ph.D. thesis, MIT, A.I. Technical Report 1442.
 101. SUETENS, P., SMETS, C., VERBEECK, G. et OOSTERLINCK, A. (1988). Knowledge-Based Blood Vessel Delineation on DSA Images. *SPIE, vol. 914, Medical Imaging II*, pp. 406-411.
 102. TENENBAUM, J.M. et BARROW, H.G. (1977). Experiments in Interpretation-Guided Segmentation. *Artificial Intelligence*, vol. 8, pp. 241-274.
 103. TERZOPOULOS, D. (1986a). Regularization of Inverse Visual Problems Involving discontinuities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 129-139.
 104. TERZOPOULOS, D. (1986b). Image Analysis Using Multigrid Relaxation Methods. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 413-424.
 105. TERZOPOULOS, D. (1987a). On matching deformable models to images. In *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, DC: 160-163.
 106. TERZOPOULOS, D. (1987b). Matching deformable models to images: direct and iterative solutions. In *Topical Meeting on Machine Vision, Technical Digest Series*, Washington, DC, pp. 164-167.
 107. TERZOPOULOS, D., PLATT, J., BARR, A. et FLEISCHER, K. (1987). Elastically

- Deformable Models. *Computer Graphics*, vol. 21, no. 4, pp. 205-214.
108. TERZOPOULOS, D., WITKIN, A. et KASS, M. (1987). Symmetry-Seeking Models and 3D Object Reconstruction. *International Journal of Computer Vision*, vol. 1, pp. 211-221.
 109. TIKHONOV, A.N. et ARSENIN, V.Y. (1977). *Solutions of Ill-Posed Problems*. Washington, DC: Wunston and Sons.
 110. TORCHU, F. (1993). A Contouring Program Based on Dual Kriging Interpolation. *Engineering with Computers*, vol. 9, pp. 160-177.
 111. TORRE, V. et POGGIO, T.A. (1986). On Edge Detection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 147-163.
 112. UEDA, N. et SUZUKI, S. (1993). Learning Visual Models from Shape Contours Using Multiscale Convex/Concave Structure Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 337-352.
 113. ULLMAN, S. (1979a). The Interpretation of Structure from Motion. *Proceedings of the Royal Society London B*, vol. 203, pp. 405-426.
 114. ULLMAN, S. (1979b). *The Interpretation of Visual Motion*. The MIT Press.
 115. ULLMAN, S. (1976). Filling-in the Gaps: The Shape of Subjective Contours and a Model for Their Generation. *Biological Cybernetics*, vol. 25, pp. 1-6.
 116. VIOLA, P.A. (1995). *Alignment by Maximization of Mutual Information*, Ph.D. thesis, MIT, AIL, A.I. Technical Report 1548.
 117. WALLACE, T.P. et WINTZ, P. (1980). An efficient, three-dimensional aircraft recognition algorithm using normalized Fourier descriptors. *Computer Graphics and*

- Image Processing*, vol. 3, pp. 99-126.
118. WITKIN, A.P. (1983). Scale-Space Filtering. *Proceedings of the International Joint Conference on Artificial Intelligence*, Karlsruhe, pp. 1019-1022.
 119. WITKIN, A.P., TENENBAUM, J.M. (1983). On the Role of Structure in Vision. In *Human and Machine Vision*, Rosenfeld *et al.* editors, New York: Academic Press, pp. 481-543.
 120. WITKIN, A., TERZOPOULOS, D. et KASS, M. (1987). Signal matching through scale space. *International Journal of Computer Vision*, vol. 1, pp. 133-144.
 121. WHITTLE, P. (1963). Stochastic Processes in Several Dimensions. *Bulletin of the International Statistical Institute*, vol. 40, pp. 974-994.
 122. ZUCKER, S. W. (1987). The Diversity of Perceptual Grouping. In *Vision, Brain and Cooperative Computation*, Hanson, A., editor, London: MIT Press, pp. 231-261.
 123. YUILLE, A. et ULLMAN, S. (1989). Rigidity and Smoothness of Motion: Justifying the Smoothness Assumption in Motion Measurement. In *Image Understanding*, Ullman, S., Richards, W., editors, Ablex Publishing Corp., pp. 163-184.
 124. ZHU, S.C. et YUILLE, A.L. (1996). FORMS: A Flexible Object Recognition and Modelling System. *International Journal of Computer Vision*, vol. 20, no. 3, pp. 187-212.
 125. ZHU, S.C. et YUILLE, A.L. (1996). Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884-900.

Annexe A

Stabilisateurs fonctionnels

Soit Y l'image observée et y l'image réelle. Y est définie sur un domaine discret D_Y de N échantillons de l'image réelle y dite sous-jacente à Y . y est une fonction continue définie sur un domaine D_y . La fonction y peut être estimée à partir des données observées Y en minimisant l'erreur quadratique moyenne entre y et Y par rapport aux paramètres définissant y :

$$\epsilon^2 = \sum_{i=1}^N (y(s_i) - Y(s_i))^2 \quad (\text{A.1})$$

Les coordonnées 3-D d'un site s_i sont notées $(x_1(s_i), x_2(s_i), x_3(s_i))$. La fonction y dans l'équation (A.1) peut aussi être remplacée par la fonction de formation d'image $H(y(s))$ si y indique la réalité 3-D. La différence entre les dérivées $y_x(s_i)$ de la solution y et les dérivées spatiales $Y_x(s_i)$ des données Y peut être élevée. Par conséquent, la stabilité de la solution ne peut pas être garantie par rapport aux perturbations des données. De plus, les échantillons de Y sont souvent fragmentés et ne sont pas uniformément distribués à cause de plusieurs facteurs reliés au processus de formation d'image. Cela signifie que le problème est sous-contraint et

la solution n'est pas unique. Par la suite, le problème est dit mal posé et il nécessite une régularisation. La forme générale de régularisation d'un problème de différentiation est d'estimer la fonction y optimisant la fonctionnelle suivante:

$$U(y, Y) = \frac{1}{\lambda \sigma^2} \sum_{i=1}^N (y(s_i) - Y(s_i))^2 + \Omega(y) \quad (\text{A.2})$$

où $\Omega(y)$ est la fonctionnelle stabilisatrice. Tikhonov et Arsenin (1977) proposent une forme générale de stabilisateur pour régularisation univariable donnée par la norme pondérée de Sobolev d'ordre p :

$$\Omega(y) = \sum_{m=0}^p \int_{D_y} w_m(s) \left(\frac{d^m y(s)}{dx_1^m} \right)^2 dx_1 \quad (\text{A.3})$$

où $w_m(s)$ sont des fonctions de pondération non-négatives, continues et spécifiées d'avance.

La solution minimisant la fonction U offre un compromis entre la conformité à l'image observée Y dans le sens des moindres-carrés et le lissage de la solution relativement au stabilisateur $\Omega(y)$. Les stabilisateurs quadratiques de Tikhonov définient des méthodes linéaires de régularisation qui sont invariantes dans l'espace. Par conséquent, leur application aux zones de discontinuité de l'image n'est pas adéquat. Ces discontinuités visuelles jouent un rôle principal dans les différents problèmes inverses de la vision artificielle, il est donc nécessaire d'utiliser des contraintes de régularisation adaptative afin de pouvoir les préserver. Terzopoulos (1986a) propose des contraintes à continuité contrôlée dont les propriétés sont ajustées dans l'espace. Pour cela, il utilise comme stabilisateur une somme de p fonctions de splines généralisées d'ordre $m=0, \dots, p$:

$$\Omega_{p,w}(y) = \sum_{m=0}^p \int_{D_y} w_m(s) \cdot \sum_{j_1+\dots+j_3=m} \frac{m!}{j_1! \dots j_3!} \left(\frac{\partial^m y(s)}{\partial x_1^{j_1} \dots \partial x_3^{j_3}} \right)^2 ds \quad (\text{A.4})$$

où $w(s) = \{w_0(s), \dots, w_p(s)\}$ sont des fonctions de pondération non-négatives et non nécessairement continues ou pré-spécifiées. Ces fonctions de pondération dépendent généralement des données observées et peuvent donc présenter des transitions abruptes. Elles peuvent être contrôlées, par exemple, par un processus détectant les discontinuités durant la recherche de la solution. On retrouve le stabilisateur de Tikhonov dans (A.4) en fixant les fonctions $w(s)$ et en se limitant à des dérivées partielles à une variable. La forme complexe, non quadratique et non convexe de la fonctionnelle U , suggère l'utilisation de méthodes d'optimisation capables de surmonter ou d'éviter les optima locaux. Terzopoulos (1986a) suggère les méthodes d'optimisation stochastiques, comme le recuit simulé (Geman et Geman, 1985), les méthodes à non-convexité graduelle (Blake et Zisserman, 1987) ou les méthodes non-linéaires à programmation dynamique. Pour ces dernières, $w(s)$ sont fixées pour l'itération en cours, et la fonction d'énergie U est considérée convexe puis optimisée par de méthodes déterministes. $w(s)$ est par la suite recalculée à partir de la dernière solution trouvée. Cependant, la convergence de cette technique ne peut pas être assurée.

Annexe B

Contraintes géométriques

L'*organisation perceptuelle* indique la faculté du système de la vision humaine de détecter les regroupements et les structures dans les images rétinienne. L'organisation joue un rôle important dans la perception visuelle (Gordon, 1989). Le monde perceptuel se présente généralement comme une organisation d'entités distinctes et ségréguées. En effet, selon Rock (1975, 1983) nous organisons typiquement le monde que nous percevons en utilisant plus ou moins les mêmes principes suivant lesquels croyons-nous le monde est organisé en réalité.

La doctrine de l'école de Gestalt en psychologie a commencé en Allemagne dans les années 1920. Elle est fondée sur l'idée suivante: l'ensemble apporte plus d'information que la somme des informations de ses éléments constitutants⁴⁹. Cela est vrai en ce sens qu'un regroupement d'éléments fournit de l'information additionnelle s'ajoutant à la somme d'information provenant de chaque élément. Les psychologues de Gestalt ont identifié certains indices considérés comme importants dans l'organisation perceptuelle. Les indices les plus connus et les plus utilisés sont la *proximité*, la *similarité*, la *fermeture*, la *continuité*, la *symétrie* et la *séparation de la figure de l'arrière plan*. Un autre indice de regroupement est celui de la *tendance commune*: par exemple, des points se déplaçant à la même vitesse et

⁴⁹

"The whole is greater than the sum of its parts."

dans une même direction sont regroupés ensemble comme faisant partie d'une même structure en mouvement (McCafferty, 1993). Ces indices de regroupement sont aussi appelés les règles de Gestalt pour l'organisation perceptuelle ou de *Pragnanz* signifiant en allemand la *simplicité* et la *consistance des formes*.

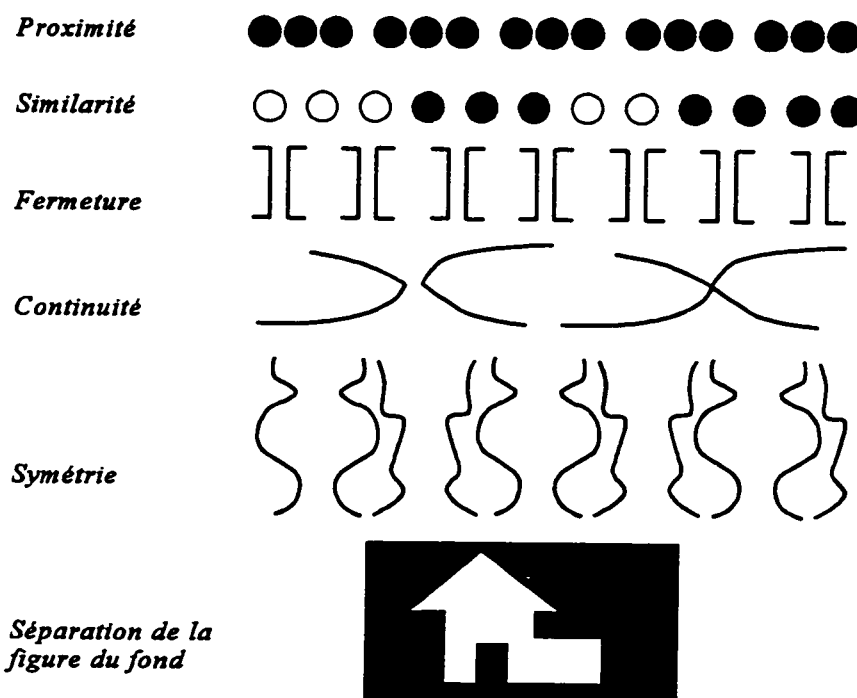


Figure B.1. Les règles de Gestalt pour l'organisation perceptuelle. Soulignons que les règles n'ont pas toutes la même importance. En cas de conflit, le regroupement se fait suivant la règle la plus dominante. Dans l'expérience de la symétrie, vérifiez que le regroupement se fait entre des lignes symétriques et non entre des lignes voisines.

Rock (1975, 1983) a suggéré que le *Pragnanz* peut être relié à la théorie de l'information en ce sens qu'un regroupement se fait entre éléments demandant la plus faible quantité de description⁵⁰. Par exemple, la symétrie, la continuité et la fermeture peuvent être utilisées pour

⁵⁰ "Minimum description length."

réduire l'information requise pour encoder la forme. Witkin et Tenenbaum (1983) ont proposé la notion de la non accidentalité pour expliquer le *Pragnanz*: les formes perçues sont celles qui ont la *vraisemblance* minimale de survenir accidentellement.

En terme de processus de traitement, l'organisation perceptuelle peut être utilisée comme une première étape de regroupement pour initialiser des procédures de reconnaissance d'objet et pour réduire l'espace de toutes les interprétations possibles (Mohan et Nevatia, 1992; McCafferty, 1993; Sarkar et Boyer, 1993; Zucker, 1987). Les facteurs de *Pragnanz* se traduisent généralement par des contraintes géométriques qui limitent significativement l'explosion de l'espace de recherche (Zucker, 1987; McCafferty, 1993; Sarkar et Boyer, 1993). Ces contraintes sont utilisées pour regrouper et organiser les traits fragmentés extraits de l'image ainsi que pour mesurer la correspondance entre les traits de l'image et ceux d'un modèle connu *a priori* (Grimson, 1990; Lowe, 1987) ou d'une autre image de la même scène comme c'est le cas pour l'inférence à partir de la stéréo-correspondance ou du mouvement (Horaud et Skordas, 1989).

Annexe C

Modèles élastiques

La théorie de l'élasticité implique l'analyse de déformation des objets. Les équations dynamiques régissant le mouvement des objets déformables peuvent être obtenues à partir de la mécanique classique de Newton. L'équilibre est atteint par l'égalité entre les forces externes appliquées à l'objet et les forces internes créées par la déformation de la matière. Soit $a=[a_1, a_2, a_3]$ le vecteur qui indique la position du point a dans un corps élastique Ω par rapport à une référence intrinsèque au corps. Les points du corps Ω sont aussi donnés dans l'espace Euclidien 3-D par les vecteurs $r(a, t)=[r_1(a, t), r_2(a, t), r_3(a, t)]$. $r(a, t)$ indique la position de la particule a à l'instant t . Ω^0 désigne le corps dans son état de repos. Un point du corps dans l'état de repos est noté $r^0(a, t)=[r_1^0(a, t), r_2^0(a, t), r_3^0(a, t)]$. Le mouvement du corps déformable est régi par les équations de Lagrange qui sont données par la forme suivante (Terzopoulos, *et al.*, 1987):

$$\frac{\partial}{\partial t} \left(\mu \frac{\partial r}{\partial t} \right) + \gamma \frac{\partial r}{\partial t} + \frac{\delta E(r)}{\delta r} = f(r, t) \quad (C.1)$$

$\mu(a)$ désigne la densité de masse du corps au point a , $\gamma(a)$ désigne le coefficient local

d'amortissement et $f(r, t)$ indique la force externe appliquée au point α à l'instant t . La fonction $E(r)$ mesure l'énergie potentielle locale créée par la déformation élastique du corps.

Le premier terme de (C.1) est la force reliée à l'inertie de masse du corps. Le deuxième terme est la force d'amortissement représentant la dissipation énergétique. Le troisième terme est la force élastique reliée à la déformation du corps ou sa déviation de son état de repos. L'énergie de déformation mesure la quantité de cette déviation et est connue aussi par l'énergie du stress subit par le corps. $E(r)$ est donné par:

$$E(r) = \int_{\Omega} \|G - G^0\|^2 da_1 da_2 da_3 \quad (C.2)$$

où G^0 et G indiquent respectivement les tenseurs métriques au point α du corps de l'état de repos Ω^0 et l'état $\Omega(t)$ à l'instant t . Le tenseur métrique 3×3 est donné par:

$$G_{ij}(r(a)) = \frac{\partial r}{\partial a_i} \cdot \frac{\partial r}{\partial a_j} \quad (C.3)$$

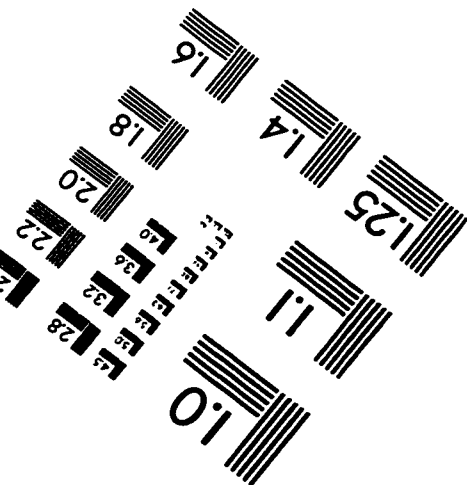
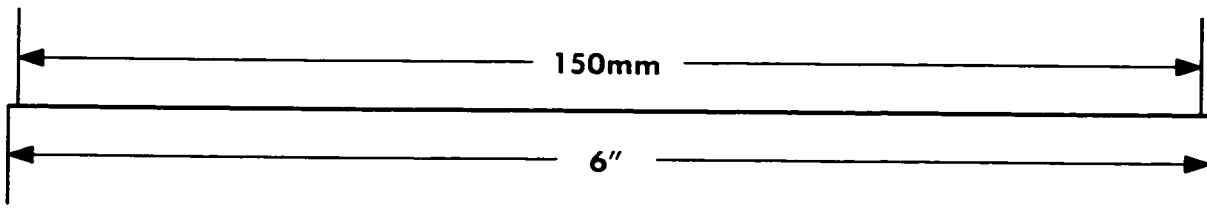
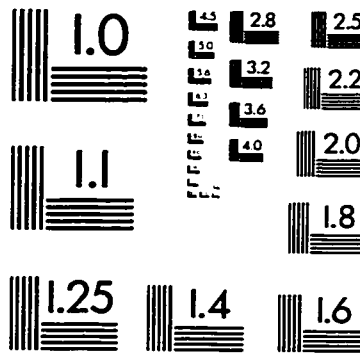
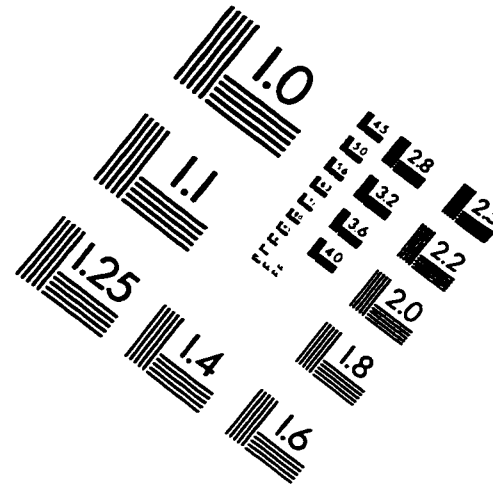
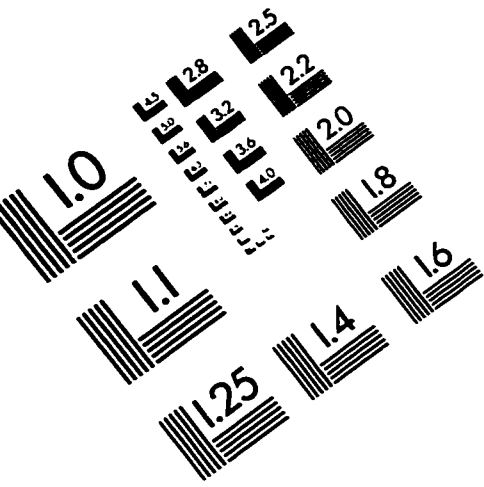
Le système d'équation (C.1) peut aussi s'écrire sous une forme discrète. Si le corps est défini sur une grille discrète de n noeuds, l'ensemble des équations qui régissent la dynamique du système est donné par l'équation suivante (Pentland et Horowitz, 1991; Terzopoulos *et al.*, 1987):

$$M \frac{\partial^2 r}{\partial t^2} + C \frac{\partial r}{\partial t} + K(r)r = f(r, t) \quad (C.4)$$

où M , C , et K sont respectivement des matrices $3n$ par $3n$ décrivant la distribution de la masse, de l'amortissement et de l'élasticité de la matière. r et $f(r, t)$ sont deux vecteurs

$3n \times 1$ indiquant respectivement la position des noeuds de la grille et l'ensemble des forces externes appliquées à chaque noeud. $f(r,t)$ est la force qui mesure l'interaction entre le modèle et l'image observée. La solution d'équilibre est trouvée par intégration temporelle du système d'équation (C.4).

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

